EXCLI Journal 2018;17:688-708 – ISSN 1611-2156
Received: June 08, 2018, accepted: July 10, 2018, published: July 20, 2018

Review article:

TOWARDS UNDERSTANDING AROMATASE INHIBITORY ACTIVITY VIA QSAR MODELING

Watshara Shoombuatong, Nalini Schaduangrat, Chanin Nantasenamat*

Center of Data Mining and Biomedical Informatics, Faculty of Medical Technology, Mahidol University, Bangkok 10700, Thailand

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

http://dx.doi.org/10.17179/excli2018-1417

ABSTRACT

Aromatase is a rate-limiting enzyme for estrogen biosynthesis that is overproduced in breast cancer tissue. To block the growth of breast tumors, aromatase inhibitors (AIs) are employed to bind and inhibit aromatase in order to lower the amount of estrogen produced in the body. Although a number of synthetic aromatase inhibitors have been released for clinical use in the treatment of hormone-receptor positive breast cancer, these inhibitors may lead to undesirable side effects (e.g. increased rash, diarrhea and vomiting; effects on the bone, brain and heart) and therefore, the search for novel AIs continues. Over the past decades, there has been an intense effort in employing medicinal chemistry and quantitative structure-activity relationship (QSAR) to shed light on the mechanistic basis of aromatase inhibition. To the best of our knowledge, this article constitutes the first comprehensive review of all QSAR studies of both steroidal and non-steroidal AIs that have been published in the field. Herein, we summarize the experimental setup of these studies as well as summarizing the key features that are pertinent for robust aromatase inhibition.

Keywords: aromatase, aromatase inhibitors, breast cancer, estrogen, QSAR, structure-activity relationship, data mining

INTRODUCTION

Breast cancer is one of the leading causes of death worldwide, with a greater prevalence in developed countries and a rapidly growing health concern in developing countries. It is also the most frequently occurring cancer found in women with an estimated 1.5 million new cases resulting in 570,000 deaths in 2015 (WHO, 2015). In addition, the prevalence of breast cancer in Asia is the highest among the world population (59 % of world population), out of which new cases account for 39 % with 44 % of cases resulting in deaths. In comparison, the prevalence of breast cancer in the continents of North America and Africa represent 5 % and 15 % of the world statistics, respectively (American Cancer Society, 2015).

Estrogen is the primary female sex hormone that acts as a double-edged sword where on one side it regulates important physiological functions for sustaining life (i.e. regulating the menstrual cycle, modulating bone density, maintenance of vessels and skin etc.) while on the other side, it is implicated in the development of breast cancer. Estrogen biosynthesis is catalyzed by aromatase, which converts androstenedione, a 19-carbon (C_{19}) steroid hormone, to estrone (E1) via a three-

Received: June 08, 2018, accepted: July 10, 2018, published: July 20, 2018

step A-ring aromatization. Aromatase also catalyzes the oxidation of testosterone, which is also then converted to estradiol (E2) (Ahmad and Shagufta, 2015) (Figure 1).

A common treatment for early-stage, hormone-sensitive breast cancer is surgery followed by radiotherapy. Furthermore, adjuvant endocrine therapy is given with or without chemotherapy depending on the tumor stage. In pre-menopausal women, most of the estrogen are made in the ovaries with the uptake of androstenedione from the circulation (Nelson and Bulun, 2001). Ovaries can convert androstenedione to estrone via the catalytic activity of aromatase, which is then transported to breast cells. However, in post-menopausal women, the main site of estrogen production are the breasts. As for the latter, the level of estrogens produced in the breast are comparable to that produced in the ovaries by pre-menopausal women, which is four to six times higher than those found in serum.

Approximately 60 % of pre-menopausal and 75 % of post-menopausal cancers are hormone-dependent (Russo et al., 2003), implying that endogenous estrogens are essentially required for proliferation. Many drugs that are used for the treatment of estrogen receptor-positive breast cancer are mechanistically based on the interference of either the estrogen production or the estrogen action.

Aromatase, also known as estrogen synthase or CYP19A1, is part of the cytochrome P450 family. It is consisting of 503 amino-acid residues spanning twelve α-helices and ten β-strands, inside which a heme cofactor is coordinated by a cysteine residue at position 437 (Ghosh et al., 2009). Aromatase is the major producer of estrogen in post-menopausal women and it catalyzes the rate-limiting step for converting androgens to estrogens (Simpson, 1994). As aromatase catalyzes the biosynthesis of estrogen from androgens, thus the inhibition of aromatase activity has be-

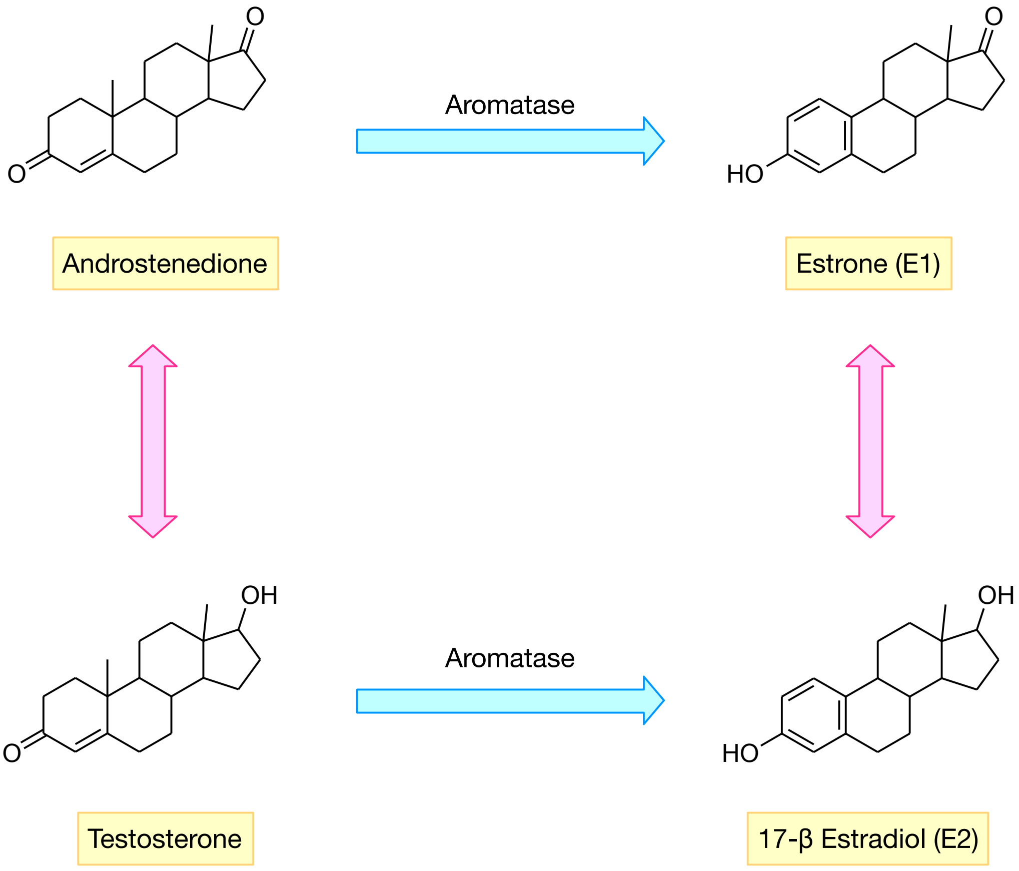

Figure 1: Summary of estrogen biosynthesis pathway as mediated by aromatase.

Received: June 08, 2018, accepted: July 10, 2018, published: July 20, 2018

come the standard treatment for hormone-dependent breast cancers in women (Eisen et al., 2008). Aromatase is located in the plasma membrane of the endoplasmic reticulum of estrogen producing cells and plays a role in development, reproduction, sexual differentiation and behavior as well as in bone and lipid metabolism, brain functions and diseases such as breast and testicular tumors. Hence, in order to block the estrogen production, it is necessary to inhibit the aromatase enzyme that is responsible for its synthesis by using aromatase inhibitors (AIs). AIs constitute the front-line therapy for estrogen-dependent breast cancers. For this reason, inhibiting this terminal step in the estradiol biosynthesis pathway is considered to be a specific and therefore, a preferable strategy.

AROMATASE INHIBITORS

To date, there are three generations of FDA-approved AIs available for inhibiting the activity of aromatase. The first-generation of AIs includes aminoglutethimide, which is marketed in the late 1970s (Santen et al., 1978, 1982; Santen and Misbin, 1981; Graves and Salhanick, 1979) (Figure 2), a derivative of the sedative agent glutethimide that was initially introduced as an anticonvulsant. However, due to its adverse effects, such as high toxicity and lack of selectivity (Demers et al., 1987; Hughes and Burley, 1970), this AI was found to interfere with other CYP450 enzymes involved in cortisol and aldosterone biosynthesis (Santen et al., 1980). Thus, aminoglutethimide was withdrawn from the market. In addition, testolactone was the first-generation steroidal AI that was used to treat advanced-stage breast cancers, albeit with weak potency (Avendaño and Menéndez, 2008). Nevertheless, these first-generation AIs served as the prototype for future generations with an emphasis on developing more potent drugs with higher selectivity and reduced toxicity. Continuing on to the second-generation, fadrozole, which contains an imidazole group (Bonnefoi et al., 1996), is more selective and potent than aminoglutethimide. Nevertheless, it still displayed effects on aldosterone, progesterone and corticosterone biosynthesis. Formestane (Brueggemeier et al., 2005), a steroidal analogue, was the first selective AI to reach clinical trials in the 1990s. It was demonstrated to be effective and was well tolerated (Dowsett and Lloyd, 1990). However, formestane exhibited poor oral bioactivity and had to be administered bi-weekly and thus, lost popularity with the discovery of the newer, more effective third-generation AIs (DrugBank, 2013).

Finally, the third-generation of AIs, are comprised of triazole derivatives, anastrozole and letrozole and one steroidal analogue, exemestane (Dutta and Pant, 2008). These AIs displayed improved efficacy and lower toxicity as compared with the estrogen antagonist, tamoxifen, in both early and advanced breast cancer (Thürlimann et al., 2004). For this reason, the last generation of AIs has been recommended by the FDA as first-line drugs for the therapy of breast carcinoma. Anastrozole and letrozole, are non-steroidal derivatives and competitive inhibitors of androstenedione. Both contain a triazole group that can interact with the prosthetic heme group of aromatase. Exemestane is a steroidal analog of androstenedione thus, permanently binding to the enzyme and deactivating its catalytic activity (Coombes et al., 2007).

Initial attempts to clarify the interaction mechanism of aromatase and its inhibitors have relied on the use of homology-derived models (Loge et al., 2005). Such studies have focused on clinically used AIs such as fadrozole, letrozole and exemestane as well as other natural products such as ligands, flavonoids and coumestrol (Karkola and Wähälä, 2009; Paoletta et al., 2008; Awasthi et al., 2015; Worachartcheewan et al., 2014b; Nantasenamat et al., 2014).

Since the crystal structure of human placental aromatase has been solved by Ghosh et al. (2009), the availability of structural details on the active site of aromatase helps in understanding of the binding characteristics of AIs as well as the evaluation of key reactions needed in the mechanism of aromatase.

EXCLI Journal 2018;17:688-708 – ISSN 1611-2156
Received: June 08, 2018, accepted: July 10, 2018, published: July 20, 2018

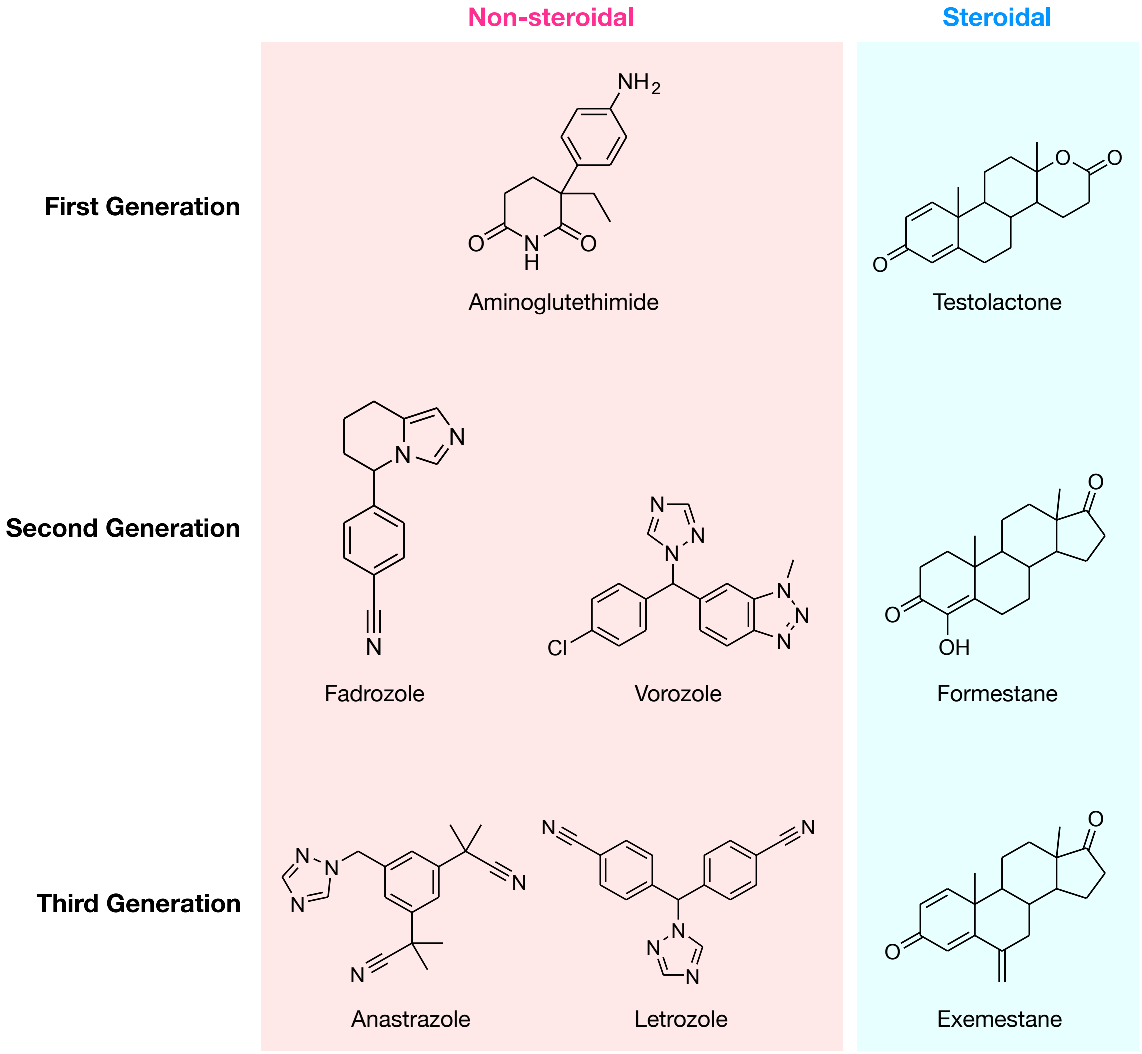

Figure 2: Chemical structures of the three generations of FDA-approved aromatase inhibitors.

This has opened up a plethora of opportunities by enabling the understanding of the molecular basis for the specificity of the aromatase enzyme and its unique catalytic mechanisms, which is imperative for the development of the next-generation of AIs.

CONCEPTS OF QSAR MODELING

Quantitative structure-activity relationship (QSAR) (Nantasenamat et al. 2009, 2010) is a ligand-based approach that seeks to discern the mathematical relationship between chemical structures (i.e. as described by various types of molecular descriptors) and the investigated biological activity through the use of statistical and machine learning techniques.

Historically, the work of Cros (1863), Crum Brown and Fraser (1868) and especially that of Muir et al. (1949) laid the foundations for the subsequent birth of QSAR as formally introduced by Hansch and Fujita (1964) in their landmark work investigating substituent effects of various compounds against various biological activities (i.e. benzoic acids against mosquito larvae, phenols against gram-positive and gram-negative bacteria, phenyl ethyl phosphate insecticides against houseflies, thyroxine derivatives against rodents, diethylaminoethyl benzoates against guinea pigs and carcinogenic compounds against mice) by utilizing substituent constants as descriptors. Ever since, QSAR had been an integral part of computational drug discovery efforts as it had been utilized to probe the underlying mechanistic basis of various biological activities (Nantasenamat and Prachayasittikul, 2015). Recently, Fujita and Winkler (2016) had shared their perspectives on the two QSAR worlds consisting of (i) classical QSAR and (ii) modern QSAR.

The early years of *classical QSAR* entails investigation on the structure-activity relationship of a congeneric set of compounds (i.e. compounds sharing a common chemical scaffold or chemotype) through the use of a few molecular descriptors. Classical QSAR methodology (Hansch et al., 1963) assumes that the biological activity of investigated chemicals can be explained by simple and interpretable physicochemical properties. These physicochemical properties encode structural features that are considered to be statistically important and that can provide useful insights and understanding pertaining to the interaction being studied. Typically, classical QSAR models are built using partial least-squares (PLS) and multiple linear regression (MLR). It should be noted that this approach does not take into consideration the 3D structure of the receptor-ligand interaction. Thus, this had inspired the development of a 3D-QSAR technique by Cramer et al. (1988) that essentially involves the alignment of a congeneric set of compounds (i.e. compounds sharing a common scaffold or chemotype) and followed by the computation of molecular fields (steric and electrostatic). Furthermore, modifications to the CoMFA concept known as comparative molecular similarity indices analysis (CoMSIA) was proposed by Klebe et al. (1994) to extend CoMFA via the utilization of Gaussian potentials as the basis for calculating similarity and thus, expand its applicability (Kubinyi, 1997).

Over the years, advancements in computation has given rise to *modern QSAR* in which an extensive list of molecular descriptors as well as a wide range of machine learning algorithms can be applied in studying the structure-activity relationship of large sets of heterogeneous and chemically diverse set of compounds. On one end, modern QSAR makes it possible to harness the big data of bioactivity information accumulated over the years for model development while on the other end, the resulting models are often complex and not readily comprehensible to bench scientists. The need for simple and interpretable QSAR models along with best practices has been discussed in a recent book chapter (Shoombuatong et al., 2017). Briefly, desirable characteristics of robust QSAR models have been set forth by the Organisation for Economic Co-operation and Development (OECD) as to encourage the utilization of QSAR models for regulatory purposes. These main OECD principles for the development of robust QSAR models are summarized in Table 1.

The typical workflow for the development of QSAR models is depicted in Figure 3. First, the QSAR modeling process starts by the compilation of a data set that entails collecting information pertaining to the compound name along with their SMILES notation, bioactivity values (e.g. IC_{50}, EC_{50}, K_i, % activity, etc.) as well as reference to the original paper. Second, the data set is subjected to data pre-processing as to ensure the completeness of the data and that there are no missing information or misspellings. Third, chemical structures are drawn and subjected to structure standardization as to remove salts, ensure appropriate charge of functional moieties, select appropriate tautomeric structures, etc. Fourth, molecular descriptors are computed as to provide quantitative description of chemical structures and this is followed by feature selection as to remove useless and/or collinear variables. Fifth, the curated data set is employed for model construction via the use of machine learning algorithms and this entails data splitting, data balancing, data validation, model validation and performance assessment. Finally, the resulting model is subjected to scrutiny on the feature importance as to identify key features contributing to the origin of the biological activity. Summary and guidelines pertaining to the best practices for QSAR model development has been described previously (Tropsha, 2010).

Table 1: The OECD principle guidelines for developing and validating QSAR model.

#	OECD principle	Description
1	Defined endpoint	To ensure that all endpoint values, within a given dataset, are consistent
2	Unambiguous algorithm	To ensure the ability of transparency and reproducibility in the proposed QSAR model
3	Defined domain of applicability	To define how robust, significant and validated QSAR model could be
4	Appropriate measures of goodness-of-fit, robustness and predictivity	To simplify the overall criteria of model validation: the internal performance of a model and the predictivity or predictive power of model
5	Mechanistic interpretation	To ensure that there are assessments of the possibility of a mechanistic interpretation

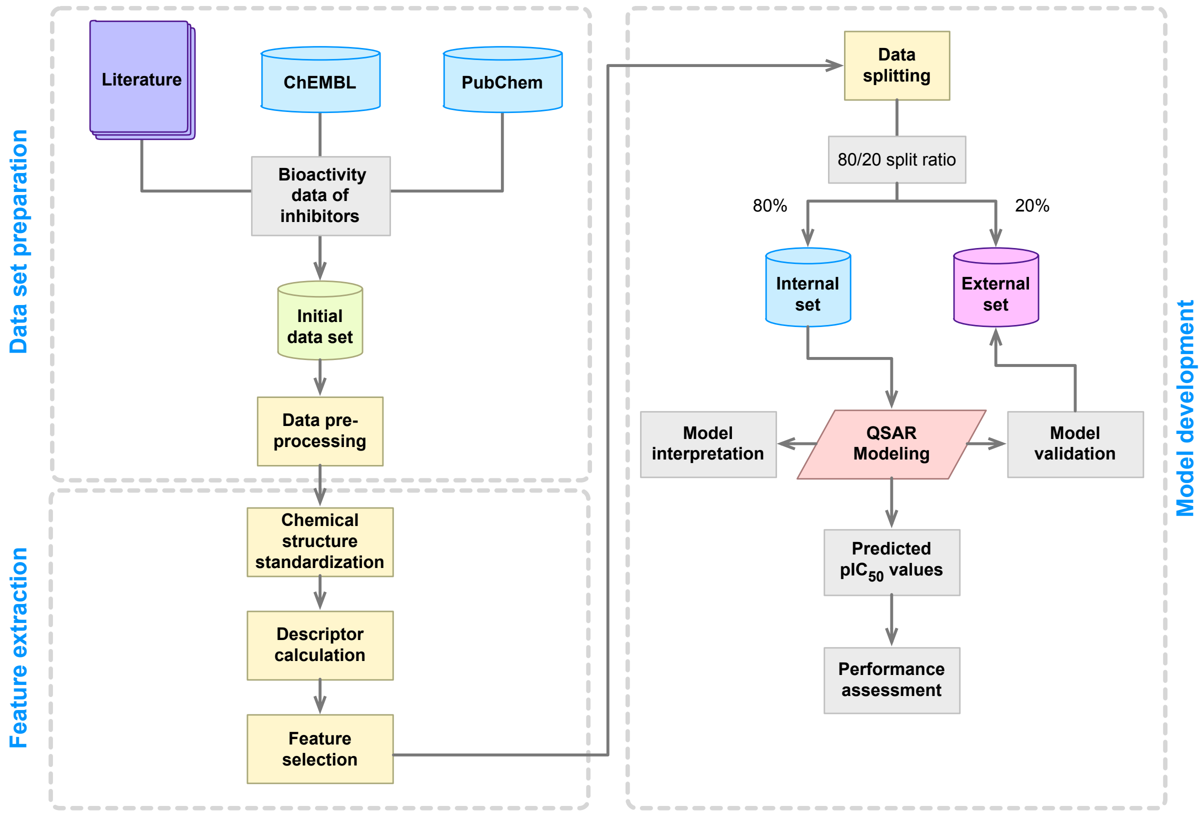

Figure 3: General workflow of QSAR model development.

Since then, the breadth of available molecular descriptors have expanded to encompass a wide range of descriptors spanning one to several dimensions. Such descriptors may account for the general features of a molecule or may consider the fine details of a molecule down to its atomic constitution. A summary of common molecular descriptors (along with its description) used in QSAR models of AIs is provided in Table 2.

Received: June 08, 2018, accepted: July 10, 2018, published: July 20, 2018

Table 2: Summary of common classes of molecular descriptors.

Descriptor class	Description	References
Molecular field	Steric and electrostatic properties of a molecule as derived from the superimposition of molecules in CoMFA analysis	Cramer et al., 1988
Molecular similarity indices	Descriptors as used in CoMSIA that are computed from Steric and Electrostatic ALignment (SEAL) similarity fields as to generate steric, electrostatic, hydrophobic, and hydrogen bonding descriptors	Kubinyi et al., 1998
Molecular surface	As implemented in CoMSA, descriptors are derived from Coulomb electrostatic potential on the molecular surface	Polanski and Gieleciak, 2003
Multivariate image analysis	Pixels derived from 2D image of chemical structures	Barigye et al., 2018
Physicochemical	Pertains to various 1D-3D chemical and physical properties of a molecule	Todeschini and Consonni, 2000
Pharmacophore mapping	A 4D-QSAR approach coupled to self-organizing map that entails the incorporation of conformational freedom into 3D-QSAR model	Bak and Polanski, 2007
Quantum chemical	Electronic properties of a molecule as derived from low-energy conformer as computed by quantum mechanical calculation	Karelson et al., 1996
SMILES	Atomic and bond constituents of a molecule	Worachartcheewan et al., 2014a
Spectral	Based on ^{13}C NMR spectroscopic data of a compound that essentially pertains to electrostatic and electronic properties as derived from frequencies of quantum mechanical properties of a nuclear magnetic moment	Beger et al., 2004

Machine learning

Machine learning is an implementation of artificial intelligence in which computers can automatically learn from data sets by extracting important patterns and making decisions or predictions. A summary of common machine learning algorithms that are used for QSAR modeling along with their strengths and weaknesses are provided in Table 3.

The concepts and in-depth treatment of machine learning is beyond the scope of this review and readers are directed to a previous comprehensive treatment of the topic (Shoombuatong et al., 2017; van Westen et al., 2011). Herein, we cover common machine learning algorithm that have been used in the study of AI activity.

The simplest learning algorithm is multiple linear regression (MLR) (Aiken et al., 2003), which is an extension of the simple linear regression and is used to explain the relationship between a series of features, X=(x_1, x_2, x_2,..., x_N), and output values, Y=(y_1, y_2, y_2,..., y_N), as follows:

$$y_i = \sum_{i=1}^{N} \beta_i x_{ij} + \beta_0 \qquad (1)$$

where y_i is the output value, x_{ij} is represented a data for i^{th} compound and j^{th} descriptor of i^{th} compound and β_i is the coefficient parameter.

Partial least squares regression (PLS). This method is a well-known method for constructing predictive models when features or descriptors have inter-correlated latent variables (Helland, 1988; Helland, 2001). PLS is closely related to principal component analysis (PCA) that consists of matrix decomposition into a matrix of eigenvectors and a matrix of its loadings factors. Given a dataset $X^{N \times M}$ with N rows and M columns, the general approach can be written as follows:

$$X^{N \times M} = X^{N \times A} P^{A \times M} + X^{N \times M} \qquad (2)$$

This is equivalent to a reduction of an *M*-dimensional variable space to an *A*-dimensional space. The variables in dimension *A* are also called latent variables.

The matrix T contains orthogonal column vectors, also called score vectors, that represents the latent variables.

Artificial neural network (ANN). This method is a computation-based method inspired by networks of biological neurons in the human brain (Puri et al., 2016). Basically, there are 3 different layers in the architecture of ANN: input layer (the input (*X*) is fed into the model through this layer), hidden layers (in general, there can be more than one hidden layers which utilizes some method to operate *X* and deliver to an output layer) and output layer (the data after processing is made available at the output layer).

Support vector machine (SVM). This statistical learning approach is based on the principles of structure-risk minimization and kernel method as proposed by Cortes and Vapnik (1995), which are used to construct a maximum-margin-separating hyperplane. The main advantage of SVM model is to seek the best compromise between the computational cost and the prediction error as to obtain the optimum generalization ability. SVM can be categorized as support vector classification (SVC) and support vector regression (SVR) (Cortes and Vapnik, 1995; Smola and Schölkopf, 2004). The principle idea of this method is to transform an input space with *m*-dimen-

Table 3: Summary of the strength and weakness of the machine-learning algorithms for performing QSAR modeling discussed in this review.

Factor	MLR	PLS	ANN	SVM	DT	RF
Non-linear			✓	✓	✓	✓
Classification and regression	✓	✓	✓	✓	✓	✓
Prediction error	High	Medium	Low	Low	Medium	Low
Computational cost	Low	Medium	Medium	Medium	Medium	High
Memory requirements	Low	Medium	Low	Low	Medium	Medium
Overfitting	✓		✓		✓	
Dimension reduction	✓	✓			✓	✓
Easy to interpret	✓	✓			✓	✓

MLR: multiple linear regression, PLS: partial least squares regression, ANN: artificial neural network, SVM: support vector machine, DT: decision tree, RF: random forest

Received: June 08, 2018, accepted: July 10, 2018, published: July 20, 2018

sional vector into a feature space with n-dimensional vector where $m < n$, and select a separating hyperplane giving the largest distance between the two classes.

Decision tree (DT). This machine learning technique is used for finding and describing a dataset (X, Y) with tree representation or structure (Safavian and Landgrebe, 1991). The tree is composed of a root node, a set of internal nodes, and a set of terminal nodes (leaves). This method is one of well-known built-in feature selector. The main purpose of using DT is to achieve a more concise and transparency of the model to identify the relationship between X and Y variables.

Random forest (RF). This ensemble learning method essentially integrates many classification and regression trees (CART). Breiman (2001) developed the RF method by growing many weak decision trees for enhancing the prediction performance of CART. The last decade has witnessed the significant achievement of RF model in applications of drug developments and related works (Win et al., 2017; Worachartcheewan et al., 2015; Pratiwi et al., 2017; Simeon et al., 2016a; Phanus-umporn et al., 2018; Suvannang et al., 2018). RF model takes advantage of two efficient machine learning techniques: bagging and random feature selection.

QSAR MODELS OF AROMATASE INHIBITORY ACTIVITY

The utilization of QSAR in aromatase research has only scratched the surface of the possible benefits that can be attained. Several classes of aromatase inhibitors have been created with only a few notable classes that have made it to the pre-clinical and clinical testing.

Thus, it is worthwhile to elucidate the physicochemical profiles of effective aromatase inhibitors in comparison with ineffective ones as such knowledge can aid in the optimization of existing compound classes or development of novel classes from available scaffolds and functional group fragments. Particularly, questions such as “What molecular descriptors are crucial for highly potent compound? How big should a potent aromatase inhibitor be? Which functional groups are most commonly found in potent compounds?” could be answered through QSAR efforts.

In 1997, Lipinski published a landmark paper on the Rule of 5 (Lipinski et al., 2001), which has been widely used in the pharmaceutical industries as general guidelines for drug development efforts. The Rule of 5 considers ADMET issues that are critical towards the success of the identified compounds of interest as it may help reduce pre-clinical and clinical failures. A similar approach may be applied to the aromatase system where several Rules may be developed for the identification of potent aromatase inhibitors.

The earliest QSAR study performed for AIs was published in 1994 by Nagy et al. (1994) whereby MLR analysis was conducted on models built with 5 quantum chemical descriptors for 24 compounds assessed by LOO-CV procedure. From the results obtained, the authors were able to discover 2 candidate AIs for further pharmacophore studies.

Furthermore, as can be seen from Figure 4 (top-left), in the years from 1994-2000 only five additional articles (Oprea and García, 1996; Recanatini, 1996; Sulea et al., 1997; Recanatini and Cavalli, 1998; Cavalli et al., 2000) on QSAR of AIs were published, which utilized mainly molecular field descriptors and LOO-CV. For example, Cavalli et al. (2000) quantitatively compared 3D-QSAR models of the cytochrome P450 active site via CoMFA modeling and homology modeling techniques. Once models were built, two nonsteroidal AIs were docked into each model and the resulting interaction energies were recorded. The authors noted that although each technique had its drawbacks, both could be used together as a mutual validation technique for ligand-based and target-based 3D models of ligand-target interactions.

In addition, Sulea et al. (1997) described van der Waals envelopes as a steric potential field in a 3D-QSAR CoMFA based modeling of ligand-receptor interactions that was performed on 78 steroidal AIs and evaluated

Received: June 08, 2018, accepted: July 10, 2018, published: July 20, 2018

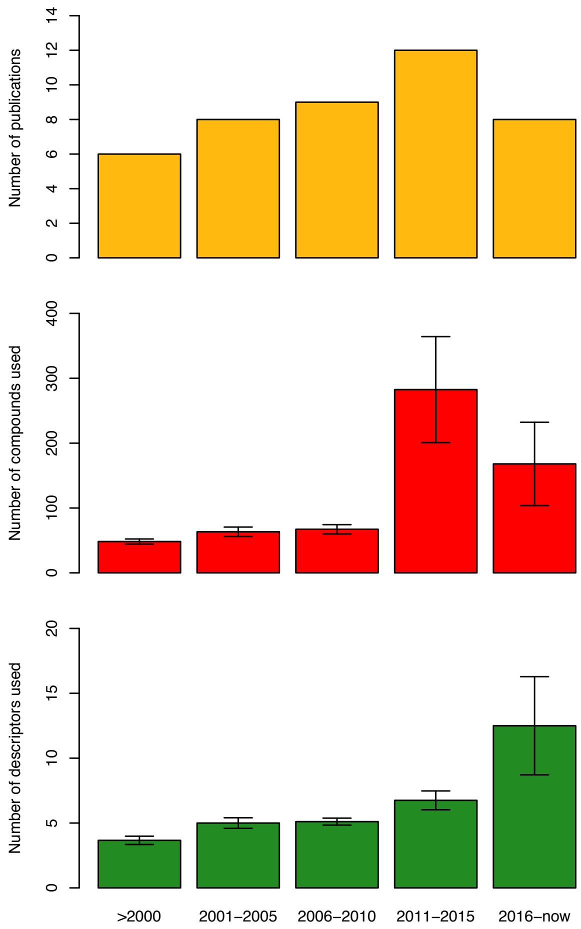

Figure 4: Overview on the number of publications (top), number of compounds (middle) and the number of descriptors (bottom) extracted from articles describing QSAR models of AIs.

LOO-CV procedure. The authors were able to prove that the van der Waals envelopes intersection volumes (INVOL) could be used as an alternative replacement for the more commonly used Lennard-Jones 6-12 potential for the identification of relevant features governing biological activities within CoMFA and 3D-QSAR based models.

Similarly, Oprea and Garcia (1996) analyzed the data of 50 steroidal AIs using CoMFA models coupled with chemometric based Generating Optimal Linear PLS Estimation (GOLPE) models validated using both the LOO-CV and the external test procedures. The authors concluded that using CoMFA, differences in aromatase inhibition among the C6-substituted steroids were shown to be consistent with known, potent inhibitors of aromatase, included in the model. In addition, when direct alignment comparisons were made, these compounds exhibited distinct features that overlapped with the steric and electrostatic fields obtained in the CoMFA model.

Over the course of the next five years (2001-2005) (Gironés and Carbó-Dorca, 2002; Beger and Wilkes, 2002; Beger et al., 2001; Polanski and Gieleciak, 2003; Beger et al., 2004; Leonetti et al., 2004; Cavalli et al., 2005), it can be seen that studies employed higher number of descriptors as well as made use of more descriptor types (e.g. molecular fields, spectral, molecular surface and quantum chemical) were observed in seven publications where the only CV method applied on the datasets was the LOO-CV (Figure 4). Beger et al. (2004) developed a technique which was similar to QSAR modeling, which they called the minimum deviation of structurally assigned spectra analysis (MiDSASA). This method was based on minimum chemical shift differences on substructure fragments instead of relying on substructure fragments as a whole for model production as is typical in SAR modeling. The authors used this MiDSASA template on 50 steroids binding the aromatase enzyme based on the average activity of the four nearest compounds, resulting in a correlation of 0.71. The authors further suggested that models made using the minimum deviation concept can be applied to other chemoinformatic data analyses such as metabolite concentrations in metabolic pathways for metabolomics research.

In addition, Beger et al. (2001) built quantitative spectroscopic data-activity relationship (QSDAR) models for 50 steroidal AIs developed based on data collected via simulated ^{13}C nuclear magnetic resonance (NMR). The models were based on comparative spectral analysis (CoSA) and comparative structurally assigned spectral analysis (CoSASA). From the PLS analysis, the CoSA models exhibited R^2 of 0.78 and Q^2 of 0.71 while the CoSASA based models provided R^2 of 0.75 and Q^2 of 0.66.

EXCLI Journal 2018;17:688-708 – ISSN 1611-2156
Received: June 08, 2018, accepted: July 10, 2018, published: July 20, 2018

Similarly, Polanski and Gieleciak (2003) used CoMSA to analyse the 3D-QSAR models built for 50 steroidal AIs. The authors aimed to predict regions that are important for the binding activity of the ligand with the enzyme. Using uniformative variable elimination as coupled to partial least squares (UVE-PLS) or modified iterative UVE procedure (IVE-PLS), the authors were able to determine that the 3D-QSAR models generated ($Q^2 = 0.96$) outperformed those reported at the time using CoMFA, CoSA or CoSASA.

Furthermore, the number of articles on QSAR of AIs were seen to increase rapidly for the years 2006-2010 (Figure 4) with the publication of ten articles in the time period (Bak and Polanski, 2007; Nagar et al., 2008; Castellano et al., 2008; Mittal et al., 2009; Gueto et al., 2009; Nagar and Saha, 2010a, b; Roy and Roy, 2010a, b; Dai et al., 2010). Most of the QSAR models in this time frame were built utilizing physicochemical descriptors as compared to other techniques in the previous years.

Additionally, the validation methods for AIs QSAR publications in the abovementioned years were tied between LOO-CV only and LOO-CV in conjunction with external validations (Figure 5). For example, Bak and Polanski (2007) conducted a 4D-QSAR study based on the self-organizing map (SOM), which is an unsupervised method based on the Kohonen neural network coupled with the IVE-PLS analysis. The use of this combined 4D-QSAR and IVE-PLS method provided a very stable and predictive modeling technique. The method enabled the authors to identify molecular motifs contributing to the aromatase enzyme binding activity. Gueto et al. (2009) employed structure-based drug design approach using receptor-independent CoMFA maps that were generated from LeapFrog calculations.

A robust model as verified by the bootstrapping method produced statistically significant results via cross-validated analysis, which consisted of 45 and 10 molecules in the training and test sets, respectively. Using this model, the authors were able to predict the activity of novel AI molecules which had more potency than previously reported compounds. Roy and Roy (2010a) performed a 3D-QSAR study on a diverse set of compounds using the crystal structure of aromatase whereby the dataset was divided into training (n=87) and testing (n=29) set by clustering techniques. All the QSAR models were subjected to multiple validation methods such as internal validation, external validation and Y-randomization. The authors concluded that in order to exhibit ideal aromatase inhibitory activity, the compound should contain at least one or two hydrogen bond acceptor groups (such as NO_2 and CN) with optimal hydrophobicity.

Additionally, the increase in popularity of QSAR models for predicting AIs was greatly observed in 2011-2015 (Narayana et al., 2012; Kishore et al., 2013; Nantasenamat et al., 2013a, b, 2014; Dai et al., 2014; Xie et al., 2014; Worachartcheewan et al., 2014a, b; Shoombuatong et al., 2015; Awasthi et al., 2015; Xie et al., 2015; Kumar et al., 2016) whereby the number of publications increased to thirteen, with an even more dramatic rise in the number of compounds used for calculating descriptors using LOO-CV and external validation (Figure 5). Worachartcheewan et al. (2014b) investigated the QSAR of coumarins as potential AIs using 7 quantum chemical descriptors. MLR was used for the analysis of models, which were shown to achieve good predictive performance as verified by LOO-CV affording Q^2 of 0.9239 and $RMSE_{CV}$ of 0.1304 while an external validation confirmed its robustness with Q^2_{Ext} of 0.7268 and $RMSE_{Ext}$ of 0.2927.

Moreover, Nantasenamat et al. (2013b) explored a set of 54 letrozole analogs as AIs in a QSAR study using MLR, ANN and SVM methods. The QSAR model was developed using a set of descriptors giving rise to important physicochemical properties (i.e. number of rings, ALogP and HOMO-LUMO) which were further used for predicting AI activity. The authors observed a strong correlation among the predicted pIC_{50} values with

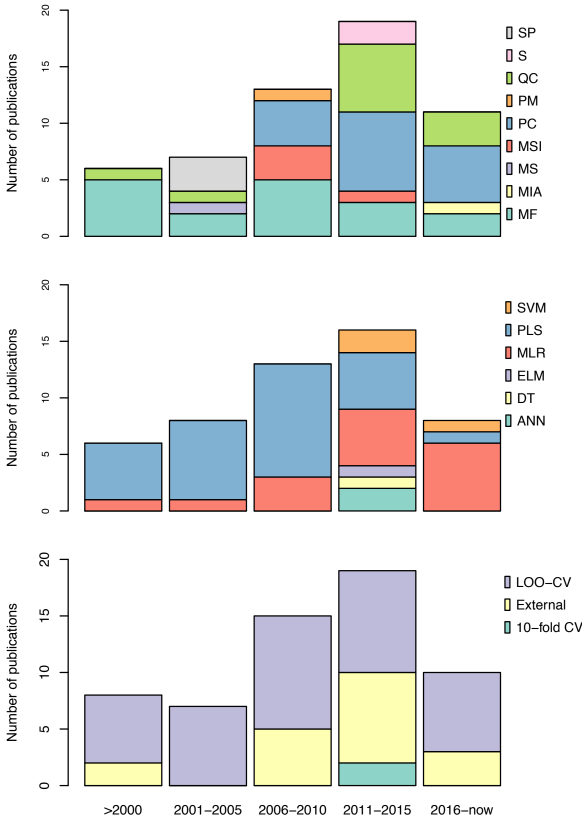

Figure 5: Overview of the types of descriptors (top), machine learning algorithms (middle) and validation methods (bottom) extracted from articles describing QSAR models of AIs.

(Abbreviations: SP, S, QC, PM, PC, MSI, MS, MIA and MF represents spectral, SMILES, quantum chemical, pharmacophore mapping, physicochemical, molecular similarity indice, molecular surface, multivariate image analysis and molecular field, respectively. SVM, PLS, MLR, ELM, DT and ANN represents support vector machine, partial least square, multiple linear regression, efficient linear model, decision tree and artificial neural network, respectively. LOO-CV, external and 10-fold CV represents leave-one-out cross-validation, external test and 10-fold cross-validation, respectively)

their experimental values, displaying correlation coefficient Q^2 values in the range of 0.72–0.83 while the external test set (Q^2_{Ext}) afforded values in the range of 0.65–0.66. Furthermore, Worachartcheewan et al. (2014a) employed the bioactivity data on pIC_{50} of 973 AIs for constructing QSAR models using CORelation And Logic (CORAL) software (http://www.insilico.eu/coral) where the molecular structures are represented by their simplified molecular input line entry system (SMILES) notation and thus eliminating the need to geometrically optimize molecular structures. The Monte Carlo optimization of correlation was used for predicting the aromatase inhibitory activity. Results obtained from rigorous dataset splits and CV techniques indicated that models were reliable with R^2 and Q^2 in ranges of 0.6271–0.7083 and 0.6218–0.7024, respectively. Similarly, Nantasenamat et al. (2014) conducted the first large-scale QSAR study on a non-redundant set of 63 flavonoids using MLR, ANN, SVM and DT methods. The models obtained showed good predictive performance with Q values in the range of 0.8014–0.9870 and 0.8966–0.9943 evaluated by LOO-CV and external test, respectively. Furthermore, in another study conducted by our group Shoombuatong et al. (2015), proposed the simple and interpretable efficient linear method (ELM) for constructing a highly predictive QSAR model. The results indicated that a robust performance was achieved using the ELM method with 10-fold CV MCC values of 0.64 and 0.56 for steroidal and non-steroidal AIs, respectively. In addition, Xie et al. (2014) constructed 3D QSAR models in order to elucidate the steroidal AIs with lower side effects using CoMFA and CoMSIA methods. The models produced were reliable and predictive good statistical results for CoMFA: $Q^2 = 0.636$, $R^2 = 0.988$, $Q^2_{Ext} = 0.658$ and CoMSIA: $Q^2 = 0.843$, $R^2 = 0.989$, $Q^2_{Ext} = 0.601$.

The current trend (2016–2018; Figure 4) shows that eight articles (Song et al., 2016; Ghodsi and Hemmateenejad, 2016; Adhikari et al., 2017a; Prachayasittikul et al., 2017; Pingaew et al., 2018; Lee and Barron, 2018; Barigye et al., 2018) have already been published in comparison to a total of 13 publications for the previous 5 years. Thus, it is promising that the number of publication regarding AIs utilizing QSAR models for prediction will continue to grow.

Received: June 08, 2018, accepted: July 10, 2018, published: July 20, 2018

To further aid in that growth process, the number of compounds used as the data set has seen a steady rise with the number of descriptors for generating QSAR models saw a dramatic increase as compared to previous years. As for the types of descriptors, the trend has moved towards modern QSAR with the utilization of physicochemical properties and quantum chemical structures to build the models. In addition, the main validation techniques remain the same as previous years whereby LOO-CV and external validation were mainly used. Ghodsi and Hemmateenejad (2016) conducted QSAR studies on a series of diarylalkylimidazole and diarylalkyltriazole derivatives previously evaluated as being potent AIs using 870 quantum chemical descriptors (such as dipole moment and energies of HOMO and LUMO orbitals, hydration energies, and lipophilicity) that were analyzed using MLR. The models were validated with the LOO-CV and the authors concluded that lipophilicity was an important factor for the strong binding to aromatase. In addition, the HOMO orbital shape and its imidazole ring distribution was also considered as important. More recently, Adhikari et al. (2017a) performed QSAR studies using various techniques (2D-QSAR, 3D-QSAR and HQSAR) on 67 non-steroidal letrozole-based analogs with promising AI activity. Stepwise multiple linear regression (SMLR) was used to build the models after which, the models were validated with the LOO for internal validation. The results from the 2D-QSAR study suggested the importance of the nitrogen atoms in their electrotopological state thereby inferring that their orientation may modulate the inhibition. The authors noted that these results were further validated with the 3D-QSAR analysis while the HQSAR model inferred the importance of the *p*-cyanophenyl moiety in regulating AI. Additionally, Lee and Barron (2018) conducted 3D-QSAR studies on the bioactivity (IC_{50}) of 124 compounds exhibiting AI activity (steroidal and heterocyclic). Multiple linear regression combined with genetic algorithm (GA-MLR) was used to build the models which was then validated via the LOO and external validation methods. Furthermore, Prachayasittikul et al. (2017) investigated the aromatase inhibitory potency of a series of 2-amino (chloro)-3-chloro-1,4-naphthoquinone derivatives by constructing QSAR models using the IC_{50} values. The models were evaluated based on MLR and LOO-CV which indicated good predictive performance (Q^2 = 0.9783 and $RMSE_{CV}$ = 0.0748) of the constructed model. Therefore, 1,4-naphthoquinone derivatives can be seen as promising compounds needed further evaluations as AIs. The most recent article published by Barigye et al. (2018) reported the first practical application of Discrete Fourier Transformation (DFT) based Multiple Image Analysis (MAI) derived 2D-QSAR model for the classification of an aforementioned set of 973 novel AIs as compiled from the literature (Nantasenamat et al., 2013a).

INSIGHTS FROM QSAR MODELS

Model interpretation is the process by which the underlying features contributing the most to the investigated biological activity are deduced as to help guide the design of novel and robust drugs. The interpretability of a QSAR model is contingent upon the types of descriptors and machine learning algorithms used. As summarized in Table 4, it can be observed that prior to 2010, MLR and PLS models, also known as white-box approaches, were the most popular and yet simple learning algorithms used for QSAR modeling of AIs.

Although these two models are interpretable but they did not perform well on highly complexed data. On the other hand, a black-box approach like ANN and SVM can provide higher accuracy in the same case but they cannot provide details pertaining to how the factors exert its influence on the biological activity of investigated compounds. Analysis of key features for aromatase inhibition from

Received: June 08, 2018, accepted: July 10, 2018, published: July 20, 2018

Table 4: Summary of machine learning algorithm used in QSAR modeling for predicting and analyzing aromatase inhibitor.

Year	Number of compounds	Type of descriptors[a]	Number of descriptors	ML algorithm[b]	Validation[c]	References
1994	24	QC	5	MLR	LOO-CV	Nagy et al., 1994
1996	29	MF	3	PLS	LOO-CV	Recanatini, 1996
1996	50	MF	2	PLS	LOO-CV, External	Oprea and García, 1996
1997	78	MF	4	PLS	LOO-CV	Sulea et al., 1997
1998	60	MF	6	PLS	LOO-CV, External	Recanatini and Cavalli, 1998
2000	49	MF	2	PLS	LOO-CV	Cavalli et al., 2000
2001	50	SP	5	PLS	LOO-CV	Beger et al., 2001
2002	50	QC	6	PLS	LOO-CV	Gironés and Carbó-Dorca, 2002
2002	50	SP	9	PLS	LOO-CV	Beger and Wilkes, 2002
2003	50	MS	5	PLS	LOO-CV	Polanski and Gieleciak, 2003
2004	35	MF	3	PLS	LOO-CV	Leonetti et al., 2004
2004	50	SP	5	MLR,PLS	LOO-CV	Beger et al., 2004
2005	70	MF	5	PLS	LOO-CV	Cavalli et al., 2005
2007	152	PM	2	PLS	LOO-CV	Bak and Polanski, 2007
2008	128	MF	5	PLS	LOO-CV, External	Castellano et al., 2008
2008	33	MF, MSI	4	MLR, PLS	LOO-CV	Nagar et al., 2008
2009	30	MF, MSI	3	PLS	LOO-CV	Mittal et al., 2009
2009	66	MF	7	PLS	LOO-CV, External	Gueto et al., 2009
2010	32	PC	7	PLS	LOO-CV	Dai et al., 2010
2010	59	PC	5	PLS	LOO-CV	Roy and Roy, 2010b
2010	116	PC	4	PLS	LOO-CV, External	Roy and Roy, 2010a
2010	52	MF, MSI	5	MLR, PLS	LOO-CV, External	Nagar and Saha, 2010a
2010	89	PC	6	MLR, PLS	LOO-CV, External	Nagar and Saha, 2010b
2012	39	PC	3	MLR	LOO-CV, External	Narayana et al., 2012
2013	54	PC, QC	3	MLR, ANN, SVM	LOO-CV, External	Nantasenamat et al., 2013b
2013	973	PC, QC	13	DT	10-fold CV	Nantasenamat et al., 2013a
2013	73	QC	5	PLS	External	Kishore et al., 2013
2014	34	PC, QC	7	MLR	LOO-CV, External	Worachartcheewan et al., 2014b
2014	973	S	7	MLR	LOO-CV, External	Worachartcheewan et al., 2014a
2014	63	PC, QC	6	MLR, ANN, SVM	LOO-CV	Nantasenamat et al., 2014
2014	14	PC	5	PLS	LOO-CV	Dai et al., 2014
2015	45	MF	6	PLS	LOO-CV, External	Awasthi et al., 2015
2015	84	MF, MSI	7	PLS	LOO-CV, External	Xie et al., 2015

Received: June 08, 2018, accepted: July 10, 2018, published: July 20, 2018

Year	Number of compounds	Type of descriptors[a]	Number of descriptors	ML algorithm[b]	Validation[c]	References
2015	973	PC, QC	15	ELM	10-fold CV	Shoombuatong et al., 2015
2015	66	MF, S	4	PLS	LOO-CV, External	Xie et al., 2014
2016	46	MF	5	PLS	LOO-CV	Kumar et al., 2016
2016	76	PC, QC	9	MLR	LOO-CV	Ghodsi and Hemmateenejad, 2016
2016	13	PC	4	MLR	LOO-CV	Song et al., 2016
2017	11	PC, QC	4	MLR	LOO-CV	Prachayasittikul et al., 2017
2017	67	MF	5	MLR	LOO-CV, External	Adhikari et al., 2017a
2018	124	PC, QC	9	MLR	LOO-CV, External	Lee and Barron, 2018
2018	34	PC	4	MLR	LOO-CV	Pingaew et al., 2018
2018	973	MIA	60	SVM	External	Barigye et al., 2018

[a]MF: molecular field, MIA: Multivariate image analysis, MS: Molecular surface, MSI: molecular similarity indices, PC: physicochemical, PM: Pharmacophore mapping, QC: quantum chemical, S: SMILES, SP: Spectral

[b]ANN: artificial neural network, MLR: multiple linear regression, PLS: partial least squares regression, SVM: support vector machine

[c]LOO-CV: leave-one-out cross-validation, 10-fold CV: 10-fold cross-validation

selected QSAR works employing descriptors pertaining to quantum chemical and physicochemical properties are performed hereafter (Table 5). Nantasenamat et al. (2013a) performed a large-scale QSAR modeling of a set of steroidal and non-steroidal AIs and revealed that the most important features from PCA analysis were found to be nHAcc, TPSA and LUMO for non-steroidal and Q_m, TPSA and nHAcc and ALogP for steroidal AIs. In addition, fragment analysis provided complementary insights by suggesting that the presence of the azole ring in non-steroidal inhibitors (i.e. known to coordinate with the heme iron) and the presence of carbonyl group in the C3 position of steroidal inhibitors were important for aromatase inhibition.

In addition, using the same set of data, Shoombuatong et al. (2015) used the ELM model to deduce the most important features associated with AI. It was observed that the top four most informative descriptors for the steroidal dataset were C-025 (atom centered fragments; R--CR--R), ESpm14u and ESpm13r (connectivity or bonding between atoms) and MATS6p (involved in polarizability of molecules).

As for the non-steroidal dataset, the most important feature was determined to be molecular graph, polarizability and electronegativity of the compound. Therefore, the authors concluded that the polarizability of the compounds along with a suitable shape may be the determining factors needed for both types of AIs for reaching its intended target. Additionally, Worachartcheewan et al. (2014a) conducted a large-scale study on AIs using SMILES-based descriptors and discovered that the most notable features were the presence of cyclic rings (i.e. found in steroidal inhibitors) and the presence in the molecular structure of oxygen atoms together with double bonds that are disconnected in the structure (++++O---B2==) (i.e. analogous to the ketone groups present in the natural substrate, androstenedione) are important in increasing aromatase inhibitory activity. Furthermore, Ghodsi and Hemmateenejad (2016) conducted QSAR on AIs based on long-chained diarylalkylimidazole and diarylalkyltriazole (non-steroidal) molecule skeletons in which they determined important features to include

Table 5: Summary of key features for aromatase inhibition as deduced from QSAR modeling. Example descriptors are shown in the parenthesis.

Steroidal	**Non-steroidal**
● Number of cyclic rings ● Lipophilic ● Polar (TPSA, MATS6p)	● Nitrogen-containing descriptors (G(N···N)) ● Polarizability (HOMO, HOMO-LUMO) ● Hydrogen bond acceptors (nHAcc)

TPSA: an empirical measure of the polar surface area of a molecule, MATS6p: Moran autocorrelation of lag 6 as weighted by polarizability, G(N···N): Sum of geometric distances between N···N, HOMO: the highest energy molecular orbitals, HOMO-LUMO: the energetic difference between the HOMO and LUMO states, nHAcc: the number of hydrogen bond acceptors present in a molecule

geometrical distances of N and N atoms as well as that of O and O atoms (i.e. nitrogen atoms of azole rings as well as oxygen atoms from steroidal ketones), length of the bridge carbon chain (i.e. methylene spacer separating the azole ring and the phenol ring), number of triple bonds (i.e. triple bond in the nitrile or CN that is an integral part of FDA-approved AIs), HOMO energy (i.e. localization of HOMO orbital predominantly in the imidazolyl ring), etc. Furthermore, Nantasenamat et al. (2014) studied flavonoids with aromatase inhibitory activity, and found that active compounds were found to exhibit smaller size, higher degree of rigidity, lower polarity and charge distribution, and afforded lower electron-withdrawing tendency and higher chemical reactivity than those of the inactive class.

As for the analysis of 3D-QSAR models utilizing descriptors based on molecular fields, Castellano et al. (2008) revealed that the aligned molecules showed the presence of three major regions in which two were pertinent for aromatase inhibition (i.e. one important region afforded both electrostatic and hydrogen bonds while the second important region was occupied by the characteristic azole moiety) whereas the other region was not important for the activity. Adhikari et al. (2017a) performed an extensive study employing a wide range of QSAR models including 2D and 3D QSAR as well as molecular docking to also confirm the importance of the electrostatic property of the nitrogen-containing azole moiety, *p*-cyanophenyl moiety, *p*-nitrophenyl, hydro-phobicity as well as the appropriate size and shape of AIs were crucial for aromatase inhibitory activity. Xie et al. (2015) performed both CoMFA and CoMSIA studies and both studies further supported the importance of bulky steric groups as well as the importance of electrostatic properties pertaining to the presence of azole nitrogen atoms.

CONCLUSION

In spite of extensive research (i.e. medicinal chemistry and QSAR work) in the quest of novel and potent aromatase inhibitors, there has been only a few review articles on the topic (Adhikari et al. 2017b; Yadav et al. 2015). Briefly, Yadav et al. (2015) carried out a review focusing on molecular modeling as well as QSAR of steroidal AIs whereas Adhikari et al. (2017b) based their review on QSAR studies of non-steroidal AIs. Herein, we have performed an extensive review on the mechanistic insights of pertinent features as derived from all previous QSAR models of both steroidal and non-steroidal AIs. Moreover, this review also summarized the experimental setup of all QSAR studies such that a comparative and holistic analysis could be deduced and used for providing a glimpse on the current state-of-the-art in the field as well as serving as the basis for planning future studies to further gain insights on aromatase inhibition. For example, it is anticipated that insights gained from QSAR models alone provides one aspect where it may be beneficial to also call upon complementary methodologies such as structure-based and systems-based approaches to facilitate and augment the ligand-based QSAR approach. In fact, there have been a few studies employing a multitude of ligand, structure and systems-based approaches in studying aromatase inhibition (Simeon et al. 2016b); Suvannang et al. 2011) and future works along this line is expected to be of great benefit to the scientific community.

Conflict of interests

The authors have declared that no competing interests exist.

Acknowledgements

This work is supported by the Office of Higher Education Commission and the Thailand Research Fund (No. MRG6180226) and the New Researcher Grant (A31/2561) from Mahidol University. This work is also supported by the Center of Excellence on Medical Biotechnology (CEMB), S&T Postgraduate Education and Research Development Office (PERDO), Office of Higher Education Commission (OHEC), Thailand. Partial support from the annual budget grant (B.E. 2557-2559) of Mahidol University is also acknowledged.

REFERENCES

Adhikari N, Amin SA, Jha T, Gayen S. Integrating regression and classification-based QSARs with molecular docking analyses to explore the structure-antiaromatase activity relationships of letrozole-based analogs. Can J Chem. 2017a;95:1285–95.

Adhikari N, Amin SA, Saha A, Jha T. Combating breast cancer with non-steroidal aromatase inhibitors (NSAIs): Understanding the chemico-biological interactions through comparative SAR/QSAR study. Eur J Med Chem. 2017b;137:365–438.

Ahmad I, Shagufta. Recent developments in steroidal and nonsteroidal aromatase inhibitors for the chemoprevention of estrogen-dependent breast cancer. Eur J Med Chem. 2015;102:375–86.

Aiken LS, West SG, Pitts SC. Multiple linear regression. In: Schinka JA, Velicer WF (eds.): Handbook of psychology: Research methods in psychology, Vol. 2 (pp 483-507). Hoboken, NJ: Wiley, 2003.

American Cancer Society. Global cancer facts & figures. 3rd ed. Atlanta, GA: American Cancer Society, 2015.

Avendaño C, Menéndez JC. Anticancer drugs that inhibit hormone action. In: Avendaño C, Menéndez JC. Medicinal chemistry of anticancer drugs (pp 53-91). Amsterdam: Elsevier, 2008.

Awasthi M, Singh S, Pandey VP, Dwivedi UN. Molecular docking and 3D-QSAR-based virtual screening of flavonoids as potential aromatase inhibitors against estrogen-dependent breast cancer. J Biomol Struct Dyn. 2015;33:804–19.

Bak A, Polanski J. Modeling robust QSAR 3: SOM-4D-QSAR with iterative variable elimination IVE-PLS: application to steroid, azo dye, and benzoic acid series. J Chem Inf Model. 2007;47:1469–80.

Barigye SJ, Freitas MP, Ausina P, Zancan P, Sola-Penna M, Castillo-Garit JA. Discrete Fourier transform-based multivariate image analysis: application to modeling of aromatase inhibitory activity. ACS Comb Sci. 2018;20:75–81.

Beger RD, Buzatu DA, Wilkes JG, Lay JO. ^{13}C NMR Quantitative Spectrometric Data-Activity Relationship (QSDAR) models of steroids binding the aromatase enzyme. J Chem Inf Comput Sci. 2001;41:1360–6.

Beger RD, Wilkes JG. Comparative structural connectivity spectra analysis (CoSCoSA) models of steroids binding to the aromatase enzyme. J Mol Recognit. 2002;15:154–62.

Beger RD, Harris S, Xie Q. Models of steroid binding based on the minimum deviation of structurally assigned 13C NMR spectra analysis (MiDSASA). J Chem Inf Comput Sci. 2004;44:1489–96.

Bonnefoi HR, Smith IE, Dowsett M, Trunet PF, Houston SJ, da Luz RJ, et al. Therapeutic effects of the aromatase inhibitor fadrozole hydrochloride in advanced breast cancer. Br J Cancer. 1996;73:539–42.

Breiman L. Random forests. Mach Learn. 2001;45:5–32.

Brueggemeier RW, Hackett JC, Diaz-Cruz ES. Aromatase inhibitors in the treatment of breast cancer. Endocr Rev. 2005;26:331–45.

Castellano S, Stefancich G, Ragno R, Schewe K, Santoriello M, Caroli A, et al. CYP19 (aromatase): exploring the scaffold flexibility for novel selective inhibitors. Bioorg Med Chem. 2008;16:8349–58.

Cavalli A, Greco G, Novellino E, Recanatini M. Linking CoMFA and protein homology models of enzyme-inhibitor interactions: an application to non-steroidal aromatase inhibitors. Bioorg Med Chem. 2000;8: 2771–80.

Cavalli A, Bisi A, Bertucci C, Rosini C, Paluszcak A, Gobbi S, et al. Enantioselective nonsteroidal aromatase inhibitors identified through a multidisciplinary medicinal chemistry approach. J Med Chem. 2005;48:7282–9.

Coombes RC, Kilburn LS, Snowdon CF, Paridaens R, Coleman RE, Jones SE, et al. Survival and safety of exemestane versus tamoxifen after 2-3 years' tamoxifen treatment (Intergroup Exemestane Study): a randomised controlled trial. Lancet. 2007;369:559–70.

Cortes C, Vapnik V. Support-vector networks. Mach Learn. 1995;20:273–97.

Cramer RD, Patterson DE, Bunce JD. Comparative molecular field analysis (CoMFA). 1. Effect of shape on binding of steroids to carrier proteins. J Am Chem Soc. 1988;110:5959–67.

Cros AFA. Action de l'alcohol amyliquesur l'organisme. Strasbourg: University of Strasbourg, 1863.

Crum Brown A, Fraser TR. On the connection between chemical constitution and physiological action; with special reference to the physiological action of the salts of the ammonium bases derived from Strychnia, Brucia, Thebaia, Codeia, Morphia, and Nicotia. J Anat Physiol. 1868;2:224–42.

Dai Y, Wang Q, Zhang X, Jia S, Zheng H, Feng D, et al. Molecular docking and QSAR study on steroidal compounds as aromatase inhibitors. Eur J Med Chem. 2010;45:5612–20.

Dai Y, Xiao Y, Wang Q, Wei S, Zhang X, Ma Z, et al. Syntheses and QSAR studies of benzylimidazole derivatives and benzylcarbazole as potential aromatase inhibitors. Asian J Chem. 2014;26: 2381-8.

Demers LM, Boucher AE, Santen RJ. Aminoglutethimide therapy in breast cancer: relationship of blood levels to drug-related side effects. Clin Physiol Biochem. 1987;5:287–91.

Dowsett M, Lloyd P. Comparison of the pharmacokinetics and pharmacodynamics of unformulated and formulated 4-hydroxyandrostenedione taken orally by healthy men. Cancer Chemother Pharmacol. 1990;27: 67–71.

DrugBank. Formestane [Internet]. 2013 [cited 2018 Jun 2] Available from: https://www.drugbank.ca/drugs/DB08905.

Dutta U, Pant K. Aromatase inhibitors: past, present and future in breast cancer therapy. Med Oncol. 2008; 25:113–24.

Eisen A, Trudeau M, Shelley W, Messersmith H, Pritchard KI. Aromatase inhibitors in adjuvant therapy for hormone receptor positive breast cancer: a systematic review. Cancer Treat Rev. 2008;34:157–74.

Fujita T, Winkler DA. Understanding the roles of the "two QSARS". J Chem Inf Model. 2016;56:269–74.

Ghodsi R, Hemmateenejad B. QSAR study of diarylalkylimidazole and diarylalkyltriazole aromatase inhibitors. Med Chem Res. 2016;25:834–42.

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

Gironés X, Carbó-Dorca R. Molecular quantum similarity-based QSARs for binding affinities of several steroid sets. J Chem Inf Comput Sci. 2002;42:1185–93.

Graves PE, Salhanick HA. Stereoselective inhibition of aromatase by enantiomers of aminoglutethimide. Endocrinology. 1979;105:52–7.

Gueto C, Torres J, Vivas-Reyes R. CoMFA, LeapFrog and blind docking studies on sulfonanilide derivatives acting as selective aromatase expression regulators. Eur J Med Chem. 2009;44:3445–51.

Hansch C, Fujita T. *p* -σ-π analysis. A method for the correlation of biological activity and chemical structure. J Am Chem Soc. 1964;86:1616–26.

Hansch C, Muir RM, Fujita T, Maloney PP, Geiger F, Streich M. The correlation of biological activity of plant growth regulators and chloromycetin derivatives with Hammett constants and partition coefficients. J Am Chem Soc. 1963;85:2817–24.

Helland IS. On the structure of partial least squares regression. Comun Stat Simul C. 1988;17:581–607.

Helland IS. Some theoretical aspects of partial least squares regression. Chemometr Intell Lab Sys. 2001; 58:97–107.

Hughes SW, Burley DM. Aminoglutethimide: a "side-effect" turned to therapeutic advantage. Postgrad Med J. 1970;46:409–16.

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

Karkola S, Wähälä K. The binding of lignans, flavonoids and coumestrol to CYP450 aromatase: a molecular modelling study. Mol Cell Endocrinol. 2009;301: 235–44.

Kishore DP, Rana A, Jain UK, Rao PM. Pharmacophore-based 3D-QSAR studies of aromatase inhibitors. Asian J Chem. 2013;25: 10588-94.

Klebe G, Abraham U, Mietzner T. Molecular similarity indices in a comparative analysis (CoMSIA) of drug molecules to correlate and predict their biological activity. J Med Chem. 1994;37:4130–46.

Kubinyi H. QSAR and 3D QSAR in drug design, Part 1: Methodology. Drug Discov Today. 1997;2:457–67.

Kubinyi H, Hamprecht FA, Mietzner T. Three-dimensional quantitative similarity-activity relationships (3D QSiAR) from SEAL similarity matrices. J Med Chem. 1998;41:2553–64.

Kumar SP, Jha PC, Jasrai YT, Pandya HA. The effect of various atomic partial charge schemes to elucidate consensus activity-correlating molecular regions: a test case of diverse QSAR models. J Biomol Struct Dyn. 2016;34:540–59.

Lee S, Barron MG. 3D-QSAR study of steroidal and azaheterocyclic human aromatase inhibitors using quantitative profile of protein-ligand interactions. J Cheminform. 2018;10(1):2.

Leonetti F, Favia A, Rao A, Aliano R, Paluszcak A, Hartmann RW, et al. Design, synthesis, and 3D QSAR of novel potent and selective aromatase inhibitors. J Med Chem. 2004;47:6792–803.

Lipinski CA, Lombardo F, Dominy BW, Feeney PJ. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. Adv Drug Deliv Rev. 2001;46: 3–26.

Loge C, Le Borgne M, Marchand P, Robert J-M, Le Baut G, Palzer M, et al. Three-dimensional model of cytochrome P450 human aromatase. J Enzyme Inhib Med Chem. 2005;20:581–5.

Mittal RR, McKinnon RA, Sorich MJ. The effect of molecular fields, lattice spacing and analysis options on comfa predictive ability. QSAR Comb Sci. 2009; 28:637–44.

Muir RM, Hansch CH, Gallup AH. Growth regulation by organic compounds. Plant Physiol. 1949;24:359–66.

Nagar S, Islam MA, Das S, Mukherjee A, Saha A. Pharmacophore mapping of flavone derivatives for aromatase inhibition. Mol Divers. 2008;12:65–76.

Nagar S, Saha A. Modeling of diarylalkyl-imidazole and diarylalkyl-triazole derivatives as potent aromatase inhibitors for treatment of hormone-dependent cancer. J Comput Chem. 2010a;31:2342–53.

Nagar S, Saha A. Exploring benzcyclo derivatives as potent aromatase inhibitors using ligand-based modeling studies. Eur J Med Chem. 2010b;45:4307–15.

Nagy PI, Tokarski J, Hopfinger AJ. Molecular shape and QSAR analyses of a family of substituted dichlorodiphenyl aromatase inhibitors. J Chem Inf Comput Sci. 1994;34:1190–7.

Nantasenamat C, Isarankura-Na-Ayudhya C, Naenna T, Prachayasittikul V. A practical overview of quantitative structure-activity relationship. EXCLI J. 2009;8: 74–88.

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

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. 2013b;69:99–114.

Nantasenamat C, Worachartcheewan A, Mandi P, Monnor T, Isarankura-Na-Ayudhya C, Prachayasittikul V. QSAR modeling of aromatase inhibition by flavonoids using machine learning approaches. Chem Pap. 2014;68:697–713.

Nantasenamat C, Prachayasittikul V. Maximizing computational tools for successful drug discovery. Expert Opin Drug Discov. 2015;10:321–9.

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.

Nelson LR, Bulun SE. Estrogen production and action. J Am Acad Dermatol. 2001;45:S116-24.

Oprea TI, García AE. Three-dimensional quantitative structure-activity relationships of steroid aromatase inhibitors. J Comput Aided Mol Des. 1996;10:186–200.

Paoletta S, Steventon GB, Wildeboer D, Ehrman TM, Hylands PJ, Barlow DJ. Screening of herbal constituents for aromatase inhibitory activity. Bioorg Med Chem. 2008;16:8466–70.

Received: June 08, 2018, accepted: July 10, 2018, published: July 20, 2018

Phanus-umporn C, Shoombuatong W, Prachayasittikul V, Anuwongcharoen N, Nantasenamat C. Correction: Privileged substructures for anti-sickling activity *via* cheminformatic analysis. RSC Adv. 2018;8:8233.

Pingaew R, Mandi P, Prachayasittikul V, Prachayasittikul S, Ruchirawat S, Prachayasittikul V. Synthesis, molecular docking, and QSAR study of sulfonamide-based indoles as aromatase inhibitors. Eur J Med Chem. 2018;143:1604–15.

Polanski J, Gieleciak R. The comparative molecular surface analysis (CoMSA) with modified uniformative variable elimination-PLS (UVE-PLS) method: application to the steroids binding the aromatase enzyme. J Chem Inf Comput Sci. 2003;43:656–66.

Prachayasittikul V, Pingaew R, Worachartcheewan A, Sitthimonchai S, Nantasenamat C, Prachayasittikul S, et al. Aromatase inhibitory activity of 1,4-naphthoquinone derivatives and QSAR study. EXCLI J. 2017;16: 714–26.

Pratiwi R, Malik AA, Schaduangrat N, Prachayasittikul V, Wikberg JES, Nantasenamat C, et al. CryoProtect: A Web server for classifying antifreeze proteins from nonantifreeze proteins. J Chem. 2017;2017: 9861752.

Puri M, Solanki A, Padawer T, Tipparaju SM, Moreno WA, Pathak Y. Introduction to artificial neural network (ANN) as a predictive tool for drug design, discovery, delivery, and disposition. In: Puri M, Pathak Y, Sutariya V, Tipparaju S, Moreno W: Artificial neural network for drug design, delivery and disposition (pp 3-13). New York: Academic Press, 2016.

Recanatini M. Comparative molecular field analysis of non-steroidal aromatase inhibitors related to fadrozole. J Comput Aided Mol Des. 1996;10:74–82.

Recanatini M, Cavalli A. Comparative molecular field analysis of non-steroidal aromatase inhibitors: an extended model for two different structural classes. Bioorg Med Chem. 1998;6:377–88.

Roy PP, Roy K. Docking and 3D-QSAR studies of diverse classes of human aromatase (CYP19) inhibitors. J Mol Model. 2010a;16:1597–616.

Roy PP, Roy K. Molecular docking and QSAR studies of aromatase inhibitor androstenedione derivatives. J Pharm Pharmacol. 2010b;62:1717–28.

Russo J, Hasan Lareef M, Balogh G, Guo S, Russo IH. Estrogen and its metabolites are carcinogenic agents in human breast epithelial cells. J Steroid Biochem Mol Biol. 2003;87:1–25.

Safavian SR, Landgrebe D. A survey of decision tree classifier methodology. IEEE Trans Syst Man Cybern. 1991;21:660–74.

Santen RJ, Misbin RI. Aminoglutethimide: review of pharmacology and clinical use. Pharmacotherapy. 1981;1:95–120.

Santen RJ, Santner S, Davis B, Veldhuis J, Samojlik E, Ruby E. Aminoglutethimide inhibits extraglandular estrogen production in postmenopausal women with breast carcinoma. J Clin Endocrinol Metab. 1978;47: 1257–65.

Santen RJ, Samojlik E, Wells SA. Resistance of the ovary to blockade of aromatization with aminoglutethimide. J Clin Endocrinol Metab. 1980;51:473–7.

Santen RJ, Worgul TJ, Lipton A, Harvey H, Boucher A, Samojlik E, et al. Aminoglutethimide as treatment of postmenopausal women with advanced breast carcinoma. Ann Intern Med. 1982;96:94–101.

Shoombuatong W, Prachayasittikul V, Prachayasittikul V, Nantasenamat C. Prediction of aromatase inhibitory activity using the efficient linear method (ELM). EXCLI J. 2015;14:452–64.

Shoombuatong W, Prathipati P, Owasirikul W, Worachartcheewan A, Simeon S, Anuwongcharoen N, et al. Towards the revival of interpretable QSAR models. In: Roy K (ed.): Advances in QSAR modeling (pp 3-55). Cham: Springer International Publ., 2017.

Simeon S, Anuwongcharoen N, Shoombuatong W, Malik AA, Prachayasittikul V, Wikberg JES, et al. Probing the origins of human acetylcholinesterase inhibition via QSAR modeling and molecular docking. PeerJ 2016a;4:e2322.

Simeon S, Spjuth O, Lapins M, Nabu S, Anuwongcharoen N, Prachayasittikul V, et al. Origin of aromatase inhibitory activity via proteochemometric modeling. PeerJ. 2016b;4:e1979.

Simpson ER. Aromatase cytochrome P450, the enzyme responsible for estrogen biosynthesis. Endocr Rev. 1994;15:342–55.

Smola AJ, Schölkopf B. A tutorial on support vector regression. Stat Comput. 2004;14:199–222.

Song Z, Liu Y, Dai Z, Liu W, Zhao K, Zhang T, et al. Synthesis and aromatase inhibitory evaluation of 4-N-nitrophenyl substituted amino-4H-1,2,4-triazole derivatives. Bioorg Med Chem. 2016;24:4723–30.

Received: June 08, 2018, accepted: July 10, 2018, published: July 20, 2018

Sulea T, Oprea TI, Muresan S, Chan SL. A different method for steric field evaluation in comfa improves model robustness. J Chem Inf Comput Sci. 1997;37: 1162–70.

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

Suvannang N, Preeyanon L, Malik AA, Schaduangrat N, Shoombuatong W, Worachartcheewan A, et al. Probing the origin of estrogen receptor alpha inhibition*via* large-scale QSAR study. RSC Adv. 2018;8: 11344–56.

Thürlimann B, Hess D, Köberle D, Senn I, Ballabeni P, Pagani O, et al. Anastrozole ('Arimidex') versus tamoxifen as first-line therapy in postmenopausal women with advanced breast cancer: results of the double-blind cross-over SAKK trial 21/95--a sub-study of the TARGET (Tamoxifen or "Arimidex" Randomized Group Efficacy and Tolerability) trial. Breast Cancer Res Treat. 2004;85:247–54.

Todeschini R, Consonni V. Handbook of molecular descriptors. Weinheim: Wiley-VCH, 2000.

Tropsha A. Best practices for QSAR model development, validation, and exploitation. Mol Inform. 2010; 29:476–88.

van Westen GJP, Wegner JK, IJzerman AP, van Vlijmen HWT, Bender A. Proteochemometric modeling as a tool to design selective compounds and for extrapolating to novel targets. Med Chem Commun. 2011;2:16–30.

WHO. Breast Cancer [Internet]. Cancer 2015 [cited 2018 Jun 1] Available from: http://www.who.int/cancer/prevention/diagnosis-screening/breast-cancer/en/.

Win TS, Malik AA, Prachayasittikul V, Wikberg JES, Nantasenamat C, Shoombuatong W. HemoPred: a web server for predicting the hemolytic activity of peptides. Future Med Chem. 2017;9:275–91.

Worachartcheewan A, Mandi P, Prachayasittikul V, Toropova AP, Toropov AA, Nantasenamat C. Large-scale QSAR study of aromatase inhibitors using SMILES-based descriptors. Chemometr Intell Lab. 2014a;138:120–6.

Worachartcheewan A, Suvannang N, Prachayasittikul S, Prachayasittikul V, Nantasenamat C. Probing the origins of aromatase inhibitory activity of disubstituted coumarins via QSAR and molecular docking. EXCLI J. 2014b;13:1259–74.

Worachartcheewan A, Shoombuatong W, Pidetcha P, Nopnithipat W, Prachayasittikul V, Nantasenamat C. Predicting metabolic syndrome using the random forest method. Sci World J. 2015;2015:581501.

Xie H, Qiu K, Xie X. 3D QSAR studies, pharmacophore modeling and virtual screening on a series of steroidal aromatase inhibitors. Int J Mol Sci. 2014;15: 20927–47.

Xie H, Qiu K, Xie X. Pharmacophore modeling, virtual screening, and 3D-QSAR studies on a series of non-steroidal aromatase inhibitors. Med Chem Res. 2015; 24:1901–15.

Yadav MR, Barmade MA, Tamboli RS, Murumkar PR. Developing steroidal aromatase inhibitors-an effective armament to win the battle against breast cancer. Eur J Med Chem. 2015;105: 1–38.