Abstract:
The development of the third-generation aromatase inhibitors viz. Letrozole, Anastrozole and Exemestane as first line drugs for estrogen dependent breast cancer treatment is a momentous achievement. The inevitable prolonged use of these first line drugs to treat breast cancer, develops different resistance mechanisms in the cancer cell. The quest for new class of aromatase inhibitors is essential to overcome the plausible resistance and unwanted side effects due to chronic therapy. Considering the magnitude for the necessity of newer generation aromatase inhibitors, the present research proposal aims at the design of new series of heterocyclic rings like benzothiazole, benzimidazole, 1,3,4-thiadiazole, 1,3,4-oxadiazole and imidazo[2,1-B][1,3,4]thiadiazole derivatives as aromatase inhibitors using Ligand Based 3D QSAR Pharmacophore Modeling and Molecular docking studies. A 3-point pharmacophore with one hydrogen bond acceptor (A) and two aromatic rings (R) was selected as pharmacophore hypothesis for building the 3D QSAR model. Training set correlation with Partial Least Square factors (R²=0.99, SD=0.1265, F=470.6, P=1.033e-018) and the test set correlation (Q² = 0.7854, RMSE = 0.5284, Pearson R = 0.9111) N = 34 was chosen among the four different PLS factors generated. Letrozole and vorozole were also included in the training and test sets for the reference. The phar-
macrophore-based alignment hypothesis was done for the new series of ligands to predict the activity. Compounds 35, 40 and 50 (predicted activity 0.8697, 0.737991 and 0.865794 respectively) have shown positive findings with respect to Letrozole and Vorozole. The Docking results also gave supportive evidence for the ligand 50 with good docking score. Interestingly the S isomers had given better docking scores when compared R counterparts. Methionine 374 H-bond interaction with ligand is needed for the selectivity to aromatase active site, some of the proposed ligands also shown H-bond interaction with Threonine 310, Alanine 306 amino acid residues. This alternate binding interactions at the opening of the active site of aromatase may open the new class of aromatase inhibitors.

**Key words:** Aromatase, Letrozole, Molecular docking, Pharmacophore, Breast cancer

## INTRODUCTION

Aromatase is the enzyme which is predominantly present in woman [1]. It selectively converts androgens specifically C19 steroids with 4-ene-3-one system to estrogens by aromatizing the ring A of steroids. Estrogens in woman plays an important role in reproduction and other homeostasis functions of cell. In premenopausal women Gonadotropin releasing hormone and Follicle stimulating hormone helps in monitoring the estrogen production by negative feedback mechanism. However in postmenopausal women there is no such mechanism. This may result in imbalance in homeostasis of cell that causes breast cancer [2]. Endometriosis, Lung cancer, Hepatic cancer and other pathological conditions [3]. Around 70% of postmenopausal breast cancer patients are found to be estrogen-dependent breast cancer cases [4]. Inhibiting the aromatase enzyme is one of the important approach to treat the estrogen dependent breast cancer [5]. The Aromatase inhibitors (AIs) used as first line drug in the treatment of postmenopausal breast cancer women are Anastrazole, Exemestane, Letrozole. These drugs are also approved for early and advanced cases of breast cancer in postmenopausal women [6,7]. These AIs bind the aromatase and prevent the binding of natural substrate (androgens).

Aromatase inhibitors chronic monotherapy or their combination with other targeted agents reported developing resistance along with side effects like osteoporosis, cardiovascular complications and others [7,8]. This resistance developed by breast cancer cells against aromatase inhibitors implicates significant reasons like integral tumor insensitivity to estrogen, aromatase inhibition unresponsiveness, production of estrogenic hormones independent of aromatase, triggering of non-endocrinial signaling pathways, bypassing apoptosis pathway and selection of hormone-insensitive cellular clones during treatment [2,9,10]. Hence aromatase inhibitors resistance has been confirmed as the major obstacle to optimal therapy management.

In-silico studies are one of the effective tools to develop new drug prototypes in short period of time to contest with the pace of new drug discovery demand. Several researchers have done insilico studies using Pharmacophore modeling [11], Structure-Guided Design using crystal structure of human placental aromatase, High-Throughput Docking which facilitated the yield of promising novel AIs and insights to the aromatase active site [12,13,14,15]. The advanced models like membrane-bound all-atom molecular dynamics simulations of aromatase reveal new insights into the fluctuations of the active site, the access channel, heme-proximal cavity, and dynamic quaternary organization of ER membrane-embedded aromatase. [16,17,18].

Some of the findings reveal that the polar, aromatic and non-polar residues of aromatase active site play a vital role in interacting with AIs [4]. The presence of heme prosthetic group at the active site provides the electrons in the form of Ferric peroxide for the conversion of androgens to estrogens [19]. Park et al with the molecular simulations studies of the aromatase found that ASP-309 residue present at the active site of aromatase is critical in active site access channels [18]. Another key interaction to be mentioned is the presence of a basic nitrogen atom (imidazole or triazole) in the structure of aromatase inhibitors apically Co-ordinate the iron atom of the heme prosthetic group of the aromatase enzyme [1,20].
The present research aims at the design of new series of heterocyclic rings like benzothiazole, benzimidazole, 1,3,4-thiadiazole, 1,3,4-oxadiazole and imidazo [2,1-B][1,3,4]thiadiazole derivatives as aromatase inhibitors using Ligand Based 3D QSAR Pharmacophore Modeling and Molecular docking studies. The pharmacophore modeling was carried out using pharmacophore alignment and scoring engine (Phase). The Phase generates partial least-square (PLS) regression, which gives significant prediction model. The predicted activity of the newly designed ligand was done using the 3D QSAR pharmacophore model. The molecular docking of the newly designed ligands was performed using Maestro.

MATERIALS AND METHODS
All computational and molecular modeling studies were carried out using the software Schrodinger, LLC, USA, 2009.

2.1. Pharmacophore based 3D-QSAR Method

2.1.1. Dataset
Thirty-four molecules of substituted letrozole based analogs as aromatase inhibitors were chosen from research work of Doiron J et al. and used for QSAR analysis [21]. All the IC50 (µM) values had been obtained using the same assay method (using the P450 Inhibition Kit CYP19/MFC (BD Biosciences, Two Oak Park Bedford, MA, USA). The IC50 values of reference compounds were checked to ensure that no difference occurred between different groups. These Letrozole based analogs series showed wide variations in their structures and potency profiles with IC50 (µM) ranging from 0.002 to 49.98. The 2D structures of 34 molecules of substituted letrozol analogs (training and test), 31 newly designed ligands were drawn using ChemBioDraw Ultra version 12.0, 2010. The 3D conformers and minimization of ligands were generated using “LigPrep” incorporated in PHASE.

2.1.2. Common pharmacophore hypothesis generation:
The common pharmacophores hypothesis (CPH) were developed using Phase, version 3.1, Schrödinger, LLC, USA, 2009. The pharmacophore features were defined and identified from the built in set of common pharmacophore features in the PHASE namely hydrogen bond acceptor (A), hydrogen bond donor (D), hydrophobic group (H), negatively charged group (N), positively charged group (P), aromatic ring (R). From the generated variants, one Hydrogen bond acceptor and two Aromatic rings (ARR) were inferred based on the highly active molecule (Letrozole). These CPH were examined by scoring function to get the best alignment of the ligands which groups together similar pharmacophores according to their inter-site distances (fig.1). The quality of each alignment is measured by alignment score (RMSD in the site-point positions), Vector score (average cosine of the angles formed by corresponding pairs of vector features, ARR in the aligned structures) and volume score (overlap of van der Waals models of the non-hydrogen atoms in each pair of structures).

The alignment was measured using survival score, defined as: 
$$S=W_{site}S_{site} + W_{vec}S_{vec} + W_{vol}S_{vol} + W_{sel}S_{sel} + W_{rew}$$
where the W’s are the weights and the S’s the scores; $S_{site}$ represents alignment score, $S_{vec}$ represents vector score, $S_{vol}$ represents volume score and $S_{sel}$ represents selectivity score. $W_{sel}$ has default value of 0.0 and $W_{site}$, $W_{vec}$, $W_{vol}$, $W_{rew}$ has default values of 1.0 which are used for the hypothesis generation. $W_{rew}$ represents reward weights defined by m-1, where m is the number of actives that match the hypothesis.

2.1.3. Generating QSAR model
In order to generate a validated QSAR model for the resolve of meaningful prediction, the available data set was randomly divided into a training set of 24 molecules and a test set of 10 molecules by incorporating IC50 activity as dependent variable and chemical diversity of aligned training set ligands. The regression was done by a partial least squares (PLS) method, in which a series of models were constructed with four different PLS factors. The PLS factor 4 (# Factor) was chosen (table 1), because of the Training set correlation with Partial Least Square factors ($R^2$=0.99, SD=0.1265, $F$=470.6, $P$=1.033e-018) gave the best
overall significance of model and statistical significance. The parameters were used to evaluate the test set prediction correlation ($Q^2 = 0.7854$, RMSE = 0.5284, Pearson R = 0.9111). F is the variance ratio. Large values of F indicate a more statistically significant regression. P is significance level of variance ratio. Smaller values of P indicated a greater degree of confidence. Q-squared is the value of $Q^2$ for the predicted activities. Pearson-R is the value for the correlation between the predicted and observed activity for the test set. The 3D QSAR models that met all these criteria concurrently gave the best predictive power.

The generated results like predicted activity of the training and test molecules were evaluated. All the compounds along with their activity and predicted activities are presented in Table 2.

The graph of Phase predicted activity and Phase activity was generated for both training and test molecules. The IC50 values in the graph are represented with polychromatic markings to correlate with phase predicted activity and phase activity. Higher values (positive) on the scattered plot indicates the higher potency of the ligands. (Fig. 3 & 4).

The equations for Partial Least Square Regression analysis for training molecules were derived from the following parameters.

$m$ is the number of PLS factors in the model  
$n$ is the number of molecules in training set  
$df_1 = m + 1$ is the degrees of freedom in model  
$df_2 = n - m - 2$ is the degrees of freedom in data  
$y_i$ is observed activity for training set molecule $i$  
$\hat{y}_i$ is predicted activity for training set molecule $i$

\[ \bar{y} = \frac{1}{n} \sum_{i=1}^{n} y_i \]  
Mean observed activity

\[ \sigma_y^2 = \frac{1}{n} \sum_{i=1}^{n} (y_i - \bar{y})^2 \]  
Variance in observed activities

\[ \text{see} = \sum_{i=1}^{n} (\hat{y}_i - y_i)^2 \]  
Sum of squared errors

\[ \sigma_{\text{err}}^2 = \frac{\text{see}}{n} \]  
Variance in errors
The statistical quantities describing the test set predictions are described below:

- $s_{sy} = \sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2$  
  Variance in model

- $SD = \sqrt{\frac{sse}{df_2}}$  
  Standard deviation of regression

- $R^2 = 1 - \frac{\sigma_{\hat{y}}^2}{\sigma_y^2}$  
  R-squared; coefficient of determination

- $F = \frac{s_{sy}/df_1}{sse/df_2}$  
  F statistic; overall significance of model

- $RMSE = \sqrt{\frac{1}{n_T} \sum_{j \in T} (\hat{y}_j - y_j)^2}$  
  Root-mean-squared error

- $Q^2 = R^2(T)$  
  Root-mean-squared error

- $r = \frac{\sum_{j \in T} (y_j - \bar{y}_T)(\hat{y}_j - \bar{\hat{y}})}{\sqrt{\sum_{j \in T} (y_j - \bar{y}_T)^2(\hat{y}_j - \bar{\hat{y}})^2}}$  
  Pearson r value, Pearson correlation coefficient

**Figure 3**: 2D structures of ligands used in 3D QSAR
**Figure 4:** 2D structures of ligands used in 3D QSAR

**Figure 5:** The scattered plot of phase predicted activity and phase activity of training set

**Figure 6:** The scattered plot of phase predicted activity and phase activity of test molecules
2.2. Docking

Taking advantage of the unexplored chemo-types as aromatase inhibitors and generated hypothesis data of ligands from the QSAR results, new subclass of ligands were proposed. Docking studies were performed using Aromatase enzyme (PDB:3S7S) on maestro-Schrodinger 9.1. 34 molecules including Letrozole and Vorozole were subjected to docking on x-ray crystallographic structure of aromatase (PDB code: 3S7S). This involved stepwise functions like protein selection, protein preparation, grid generation, ligand preparation and ligand docking studies. Ligands with only good comparable docking scores are presented in table 5 as well as in Figures 5 & 6.

All the ligands structures were drawn using ChemBioDraw Ultra 12.0. The ligands were then subjected to ligprep for generating 3D optimal and minimum energy conformers. The aromatase protein having Exemestane as a substrate in its active site (PDB: 3S7S from RCSB Protein Data Bank) was selected for the docking studies. Once the protein was imported on to the maestro docking panel, the source ligand (Exemestane) in the active site was removed followed by the preprocess like optimization and minimization of energies removal water molecules etc. Then a grid is generated at the active site of aromatase protein (PDB code: 3S7S). The prepared new set of ligands were then subjected to glide for the ligand docking. The glide docked poses were minimized by local optimization features from prime. OPLS_2005 force field and GB/SA continuum solvent model were used for calculating the energies of ligand docking complexes. The scoring algorithm was then carried on energy minimized poses to generate Glide/Dock score. The each new ligand docked poses on the Aromatase active site displayed on the workspace were enhanced by using Glide XP Visualizer to display representations like important hydrophobic, pi-pi stacking interactions and hydrogen bonds between the receptor and the ligand in the Workspace.
3. RESULTS AND DISCUSSION

3.1. QSAR results

3-point pharmacophore model ARR, with one hydrogen bond acceptor (A), and Two aromatic rings (R). The generated pharmacophore-based alignment hypothesis (Q2 = 0.785, RMSE = 0.528, Pearson R = 0.911, N = 34) was used to derive a predictive activity of 3D-QSAR model for the training set and test set.

The visualized 3DQSAR model in the Workspace (Fig 2) is used to identify ligand features that contribute positively or negatively to the predicted activity. Letrozole is represented in this as prototype 3D QSAR model (Fig.2). The region of blue cubes at Cyano functional group of the ligand indicates the favorable environment for electron withdrawing groups which affects the biological activity positively. The blue cubes on hydrogen bond acceptor region also indicate favorable environment to attach hydrogen bond acceptor atoms like Oxygen and Nitrogen which affects biological activity positively. The red cubes slightly away from the H-bond acceptor region indicate unfavorable environment to attach any functional groups which affects biological activity negatively. The representation of higher vector score values indicates prominent role of the vector features like acceptors, aromatic rings in the aligned structures. The volume score is based on the overlap of van der Waals models of the non-hydrogen atoms in each pair of structures. Thus the hydrogen bond acceptor (A) like oxygen or Nitrogen is essential for increasing the potency of the ligand (ligands 35 & 50). Attaching any hydrophobic functional groups near hydrogen bond acceptor region (A) has resulted in negative predicted activity values (Table 3). The extension of hydrogen bond acceptor (A) region with imidazo[2,1-B][1,3,4]thiadiazole or imidazo[2,1-B][1,3,4]oxadiazole ring resulted in decreased potency of the ligand which can be seen with the negative predicted activity values. However attachment of benzothiazole (aromatic) in this region influenced the biological activity positively (ligand 40) because of the degree of free rotation of bond on the nitrogen unlike fused heterocyclic rings.

The proposed molecules in the 3D-QSAR results showed that ligands 35, 40 and 50 have good predicted activity (0.8697, 0.7379 and 0.86579 respectively) close to Vorozole. Letrozole showed far higher predicted activity (1.602764) as expected.

3.2. Docking results

Using Glide XP Visualizer on the maestro workspace, the ligand receptor interactions were enhanced by de-
veloping the active site surface mesh. One of the interesting finding is the aromatase receptor site shape. It looks typically like an Iodine flask. The L-phenylalanine 221, L-Valine 313 makes up the hydrophobic conical entrance to the active site. L-Threonine 310, L-Serine 478 (negatively charged amino acids in red) forms the neck region and L-Aspartic acid 309 acts as a lid, allowing or preventing the substrate entering to the active cite. Methionine 374, L-Alanine 306, Valine 369, Valine 370, L-Leucine 372, Valine 373, L-Tryptophan 224, Phenylalanine 134, L-Leucine 477 forms the hydrophobic conical body (amino acids in green). Heme prosthetic group (grey) forms the flat bottom of the flask which plays the critical role in providing electrons to the substrate for aromatization.

![Letrozole](image1)

![Vorozole](image2)

**Figure.10:** 3D representation of Ligand receptor interactions

All the ligands that fit the active site with good docking score gave the efficiency of the ligand. However other parameters like H bond interaction and interaction of the ligand with heme prosthetic group present at the active site coners the selectivity of ligand to aromatase. Hence the generated docking score was then interpreted using H-bond interactions and hetero atoms interaction with heme prosthetic group of aromatase active site, and other hydrophobic interactions with reference to the standard drugs Letrozole and Vorozole.

Ligands 67 and 45(s) (fig 10) represented bet-
ter docking score than Letrozole and Vorozole. The H

-Bond interactions of 67 and 45(s) with the amino acids at the active site varied with that of Letrozole and vorozole interactions. Letrozole and Vorozole Characteristically had H-bond interaction with MET 374 amino acid of aromatase. Heme prosthetic group of aromatase involves in most of the Pi-Pi stacking interaction with 5 membered heterocyclic aromatic functional group (like triazole ring in Letrozole) of the ligands [1,20]. The H-bond interaction of electron withdrawing group (like -C≡N in letrozole), with MET 374 or THR 310 is important for selectivity of the drug to active site [1,4,22,].
We found that S configuration of all ligands showed better docking scores compared to their R counterparts (Table 4). This indicates the S configuration of the ligands are mostly exposed to important binding interactions, that is MET 374 and Heme prosthetic group in the active site. Along with these the presence of Phenyl/aromatic functional group of ligands played an important role in hydrophobic interactions at the active site of aromatase. The docking results of R-configuration ligands presented that ARG 115, PHE 134, TRP 224 as the other important binding interactions of aromatase which are involved in a hydrophobic or Pi-Pi stacking interaction with Phenyl or aromatic functional group of the ligands.

The ligands which have H-bond interaction with an amino acid like MET 374, THR 310 also have Pi-Pi stacking interaction with heme prosthetic group as it can be observed in ligand 45(s) and 50(s). The presence of sixteen different aromatase mutants at amino acid residues Isoleucine133, Phenylalanine235, Isoleucine395, Asparagine396, Glutamic acid 302, Proline 308, Aspartic acid 309, Threonine310, Serine 478, and Histidine 480 was reported [23, 24]. These mutants dramatically increases or decreases the binding affinity of ligand with the amino acid residues at the active site of aromatase [22]. From the docking results of the research, we found that ligands like 45(s) which have primary amine functional group (-NH₂) attached to 5 membered heterocyclic ring have H-bond interaction with THR 310 and SER 478. This dual interaction increases the possibility of alternate binding interactions when the amino acid residues are mutated.

4. CONCLUSIONS

In the present study, ligand based 3D QSAR pharmacophore modeling and docking studies of non-steroidal Aromatase Inhibitors was done for a series of novel Benzothiazole, Benimidazole, 1,3,4-thiadiazole, 1,3,4-oxadiazole, imidazo[2,1-B][1,3,4] thiadiazole derivatives. 3D Pharmacophore model ARR.1 was chosen with PLS factor 4 as the best model because the Training set correlation with Partial Least Square factors gave the best overall significance of model and statistical significance and the test set prediction correlations. The aromatase inhibitor potency of proposed ligands was influenced by Hydrogen bond acceptor and electron withdrawing groups on aromatic rings which is represented in the vector score and on the 3D pharmacophore model (fig.3). Benzothiazole, 1,3,4-thiadiazole, 1,3,4-oxadiazole containing n-methyl benzimidazole derivatives exhibited good predicted activities. However imidazo[2,1-B][1,3,4]thiadiazole derivatives gave negative predicted activity values. Any fused heterocyclic ring on the hydrogen bond acceptor hindered the aromatase inhibitory potency. The most virtually potent compounds of proposed ligands are 35, 40 and 50.

The molecular docking studies were carried out on 34 molecules including Letrozole and Vorozole. The molecules with good docking score were chosen here. The activity of the molecules was analyzed based on the docking score and hydrogen bond interaction of the ligand with the receptor. The H-bond interaction of ligands with Methionine 374 represents the selectivity of ligands with the aromatase active site. The proposed ligands binding with different amino acids of receptor and differentially binding with heme prosthetic group of aromatase active site emphasizes not only selectivity but also the alternate binding interactions at the active site of aromatase when compared with Letrozole and vorozole. These alternate binding interactions may help overcome mutated amino acids irresponsiveness due to chronic therapy. These novel chemo-types could be useful in developing rationale non-steroidal molecules for the aromatase inhibitory activity.

ACKNOWLEDGMENT

The authors would like to thank the Member Secretary, Board of Research in Nuclear Science (BRNS), Government of India, for funding the Project (Sanction No.35/14/25/2017-BRNS/35240) and Dr. N. D. Jivani, President, Shree Dhanvantary Pharmacy College for providing necessary facilities.
REFERENCES