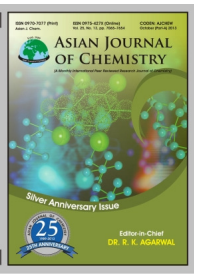
Asian Journal of Chemistry; Vol. 25, No. 18 (2013), 10588-10594

ASIAN JOURNAL OF CHEMISTRY

http://dx.doi.org/10.14233/ajchem.2013.16288



Pharmacophore-Based 3D-QSAR Studies of Aromatase Inhibitors

DEB PRAN KISHORE[1,*], AJAY RANA[2], UPENDRA KUMAR JAIN[2,*] and P. MALLIKARJUNA RAO[1]

[1]Department of Pharmaceutical Chemistry, School of Pharmacy, International Medical University, 126, Jalan Jalil Perkasa 19, Bukit Jalil 57000, Kuala Lumpur, Malaysia
[2]Pharmaceutical Chemistry Division, Chandigarh Colleges of Pharmacy, Landran, Mohali-140 307, India

*Corresponding authors: E-mail: prankishore@imu.edu.my, director@ccpmohali.org

(*Received*: 5 September 2013; *Accepted*: 28 November 2013) AJC-14453

In the present study, a ligand-based 3D-QSAR pharmacophore model have been developed by considering 73 structurally diverse classes of heterocyclic compounds (IC_{50} = 0.0077-100 μM) such as 2-phenylpyrroloquinolin-4-ones, 7-phenyl-3*H*-pyrrolo[3,2-*f*]quinolinones, 5,4'-diamino-6,8,3'-trifluoroflavones, plumbagin and N^1-(flavon-7-yl)amidrazone derivatives which were found to have potent cytotoxic activity against MCF-7 cell lines (breast cancer cell lines) to understand the structural requirements for effective binding with the aromatase enzyme to design a series of new potent and selective aromatase inhibitors (AIs). The 3D-QSAR pharmacophore hypotheses **AADRR.22** with survival score = 3.774 was found to be statistically most significant (SD = 0.4396, R^2 = 0.8679, F = 60.4 RMSE = 0.5038, Q^2 = 0.8042, Pearson-R = 0.9044) and could successfully identify known aromatase inhibitors and differentiate between active and inactive inhibitors. The potential of this model can be effectively used to design new potent aromatase inhibitors for breast cancer treatment.

Key Words: Aromatase inhibitors, MCF-7 cell lines, Breast cancer, 3D-QSAR, Pharmacophore modelling.

INTRODUCTION

Breast cancer is the most frequently diagnosed female cancer and is one of the leading cause of death in among women[1]. The hormone-dependent breast cancer require estrogens for proliferation of cell. It has been observed that interfering with endogenous hormone production by inhibiting the enzyme aromatase (responsible for conversion of androgen to estrogen) would be a promising approach for the treatment of hormone-sensetive breast cancer[2-4]. Recently, several attempts were made to design and develop potent aromatase inhibitors (AIs)[1,5-7]. Literature reveals several heterocyclic compounds such as 2-phenylpyrroloquinolin-4-ones[8], 7-phenyl-3*H*-pyrrolo-[3,2-*f*]quinolinones[9], 5,4'-diamino-6,8,3'-trifluoroflavones[10], plumbagin hydrazones[11] and N^1-(flavon-7-yl)amidrazone[12] derivatives as potent aromatase inhibitors and effective antiproliferative agents against MCF-7 (breast cancer) cell lines.The growing restrictions and costs associated with synthesis, animal experimentation, hard-pressed the researchers to look for noninvasive or virtual screening models. In pursuance of our ongoing efforts to design and develop novel non-steroidal aromatase inhibitors,[1,2] a ligand based 3D-QSAR pharmacophore model is developedby considering 73 molecules belonging to structurally diverse classes of heterocyclic compounds (IC_{50} = 0.0077-100 μM)[8-12] which were found to have potent cytotoxic activity against MCF-7 cell lines (breast cancer cell lines) to gain insight into the structural requirements for effective binding with the aromatase enzyme. The developed pharmacophoremodel **(AADRR.22)** satisfactorily reflects the structure-activity relationship of the existing aromatase inhibitors and have all the potential to facilitate the process of design and development of new potent aromatase inhibitors.

EXPERIMENTAL

Molecular modeling investigations were carried out using Dell Precision work station T3400 running Intel Core2 Duo Processor, 4GB RAM, 250 GB hard disk and NVidia Quodro FX 4500 graphics card. Maestro 9.4, Phase v3.5 pharmacophore software program (Schrodinger Inc.) was employed for the pharmacophore based 3D-QSAR studies[13,14].

Dataset: The antitumor activities of potent compounds (n = 73) such as 2-phenylpyrroloquinolin-4-ones, 7-phenyl-3*H*-pyrrolo[3,2-*f*]quinolinones, 5,4'-diamino-6,8,3'-trifluoro-flavones, plumbagin and N^1-(flavon-7-yl)amidrazone derivatives[8-12] (Table-1) against MCF-7 breast cancer cell lines have been used as a model data set to develop pharmacophore hypothesis. The biological activity data of these 73 compounds [IC_{50} (μM)] were converted to logarithmic scale [pIC_{50} (μM)] and then used for subsequent QSAR analyses as the response variable. The total dataset (n = 73) has been divided into training set (n = 52, 70 % of the total number of compounds) and test

(external evaluation) sets (n = 21, 30 % of the total number of compounds) for validation of the developed model (Table-2).

Ligand preparation: Before undertaking the task of pharmacophore model development, low-energy 3D structures must be available for each molecule of interest. Accordingly, a structure cleaning step utilizing **LigPrep** have been incorporated which attaches hydrogen, converts 2D structures to 3D, generates stereoisomer and, optionally, neutralizes charged structures or determines the most probable ionization state at user-defined pH. All the structures were ionized at neutral pH 7. Conformers for each ligand were generated using **ConfGen** by applying OPLS-2005 force field method with implicit GB/SA distance-dependent dielectric solvent model at cutoff root mean square deviation (RMSD) of 1 Å with 1,000 iterations using water as solvent.

Pharmacophore hypothesis generation: Using PHASE, one can exhaustively identify the spatial arrangements of functional groups that are common and essential for the biological activity of the ligands under investigation. The most dominating features, hydrogen bond acceptor (A), hydrogen bond donor (D), hydrophobic group (H), negatively charged group (N), positively charged group (P) and aromatic ring (R), were defined by a set of chemical structural patterns with the requirement that all five matches. Pharmacophore matching tolerance was set to 1 Å. Hypotheses were generated by a systematic variation of number of sites (n sites) and the number of matching active compounds (n act). With n act = n act - tot initially (n act - tot is the total number of active compounds in the training set), n sites. Pharmacophore models containing five sites were generated using a terminal box size of 1 Å with 25 highly active molecules (Table-2), belonging to reported aromatase inhibitors, selected using a tree based partition algorithm. To find the common pharmacophore hypothesis (CPH), the dataset was divided into active and inactive sets (Table-2). Molecules with pIC_{50} values higher than 9 were considered to be active (n = 25) and those with pIC_{50} values less than 8 were considered to be inactive (n = 25), whereas those in-between were considered to be moderately active (n = 23). Pharmacophore models containing more than five sites could not be generated and the three and four featured common pharmacophore hypothesis were rejected, as they were unable to define the complete binding space of the selected molecules. A total of 3278 probable five-featured common pharmacophore hypothesis belonging to 27 variants were identified and subjected to stringent scoring function analysis with respect to actives using default parameters for site, vector and volume. Reference relative conformational energy (kJ/mol) was included in the score with a weight of 0.01 and ligand activity, expressed as pIC_{50}, was incorporated with a weight of 1.0. A total of 110 hypotheses emerging from this process were subsequently scored with respect to the 25 inactives, using a weight of 1.

Assessment of significant hypotheses using partial least square (PLS) analyses: The generated common pharmacophore hypothesis were evaluated and validated by correlating the observed and estimated activities by randomly selecting 30 % molecules as test sets using partial least square analysis. The common pharmacophore hypothesis with the best predictive value and significant statistical data were chosen for alignment of molecules. partial least square analyses performed in which a series of models were constructed with an increasing number of partial least square factors. The accuracy of the models increases with increasing number of partial least square factors until over-fitting starts to occur either atom or pharmacophore-based model using grid spacing of 1 Å. Score hypotheses step was employed to align the actives to the hypotheses and calculate the score for the actives. common pharmacophore hypothesis of significant statistical values were selected for molecular alignments (Table-3).

Pharmacophore model development: Based on sites, 3278 common pharmacophore hypothesis with 27 variants were developed common in all active molecules (n = 25). Total five hypotheses (AAHRR.42, ADHHR.784, ADHHR.876, AADRR.22, ADHHR.660) were found to be statistically very significant. A summary of statistical data of these best 05 common pharmacophore hypothesis is listed in Table-3. Finally the hypotheses namely **AADRR.22** (survival score = 3.774), selected for molecular alignment. Partial least square analyses were conducted using five factors with a grid spacing of 1 Å. The best-fitted regression model as shown in Table-3 (SD = 0.4396, R^2 = 0.8679, F = 60.4 RMSE = 0.5038, Q^2 = 0.8042, Pearson-R = 0.9044) used for activity prediction of training and test sets molecules. Other pharmacophore models (AAHRR.42, ADHHR.784, ADHHR.876, ADHHR.660) with good statistics (Table-3) were not able to map diverse aromatase inhibitors having aromatase inhibitory activity. The distances and angles between the pharmacophoric features of the best selected pharmacophore hypothesis **(AADRR.22)** are depicted in Figs. 1 and 2 respectively. The pink ball showed hydrogen bond acceptor site (A), blue ball showed hydrogen bond donor (D), while the brown ring (R) demonstrates the ring aromatic respectively.

Alignment of ligands: In a ligand-based technique, molecules were aligned on template, using best-fitted and statistically significant common pharmacophore hypothesis **(AADRR.22)** as shown in Fig. 2a.

PHASE QSAR statistics and model validation: The PHASE descriptors served as independent variables and activity values as dependent variables in deducing 3D-QSAR models by partial least square regression analysis method. PHASE QSAR models do not use internal cross-validation techniques but rather use distinct training and test sets. Pharmacophore-based 3D-QSAR models were generated for the hypotheses **(AADRR.22)** using the 52 (70 %) member training set with five partial least square factors and a grid spacing of 1 Å. The predictive ability of each analysis was determined from a test set of 21 ligands (30 %) that were not included in the training set. The test set molecules were aligned and their activities were predicted by each partial least square analysis. In the present study, atom-based 3D-QSAR model have been developed with statistical values SD = 0.4396, R^2 = 0.8679, F = 60.4 RMSE = 0.5038, Q^2 = 0.8042, Pearson-R = 0.9044 respectively, which was significant in whole series of ligands.

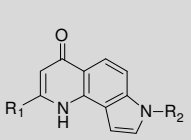
3D-QSAR Plots: The QSAR results can best be visualized using 3D plots of crucial pharmacophore regions (Fig. 2). The cubes (region) that represent the model are colored according to the sign of their coefficient values. Blue cubes refer to region

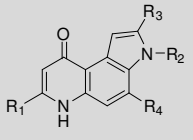
as positive coefficient, which is important for increase in activity, whereas red cubes refer to negative coefficients indicating the same ligand feature substitution with decrease in activity. The scatter plot of actual activity (Exp. pIC_{50}) *vs*. activity predicted (Pred. pIC_{50}) using PHASE is shown in Fig. 3.

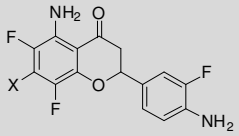
RESULTS AND DISCUSSION

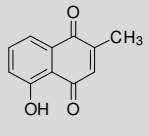
Pharmacophore modeling presents a qualitative picture of the geometry of the active site by identifying the chemical features for binding of ligands with the receptor and their spatial

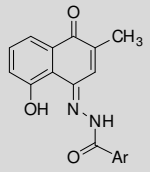
TABLE-1
DATASET USED FOR THE PHARMACOPHORE-BASED 3D-QSAR STUDIES

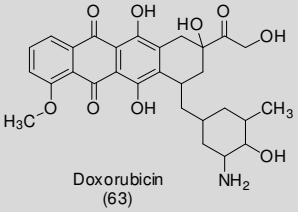

(1-9)

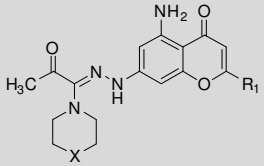

(10-22)


(23-57)


(58)


(59-62)


Doxorubicin
(63)


(64-73)

Comp. No.	R_1	R_2	R_3	R_4	Ar	X	Ref
1.	H	H	--	--	--	--	8
2.	CH_3	H	--	--	--	--	8
3.	thienyl	H	--	--	--	--	8
4.	phenyl	H	--	--	--	--	8
5.	phenyl	CH_3	--	--	--	--	8
6.	*p*-NO_2-phenyl	H	--	--	--	--	8
7.	*m*-NO_2-phenyl	H	--	--	--	--	8
8.	*m*-OCH_3-phenyl	H	--	--	--	--	8
9.	*m*-NH_2-phenyl	H	--	--	--	--	8
10.	phenyl	H	H	H	--	--	9
11.	3'-OCH_3-phenyl	H	H	H	--	--	9
12.	3'-CH_3-phenyl	H	H	H	--	--	9
13.	3'-Br-phenyl	H	H	H	--	--	9
14.	3'-CN-phenyl	H	H	H	--	--	9
15.	4'-CN-phenyl	H	H	H	--	--	9
16.	3'-NO_2-phenyl	H	H	H	--	--	9
17.	3'-NH_2-phenyl	H	H	H	--	--	9
18.	thienyl	H	H	H	--	--	9
19.	phenyl	$(CH_2)_2N(C_2H_5)_2$	H	H	--	--	9
20.	phenyl	H	$COOCH_3$	OCH_3	--	--	9
21.	CH_3	H	H	H	--	--	9
22.	H	H	H	H	--	--	9
23.	--	--	--	--	--	H	10
24.	--	--	--	--	--	CH_3	10
25.	--	--	--	--	--	C_2H_5	10
26.	--	--	--	--	--	*n*-C_4H_9	10
27.	--	--	--	--	--	*n*-C_6H_{13}	10
28.	--	--	--	--	--	Cl	10
29.	--	--	--	--	--	Br	10
30.	--	--	--	--	--	OH	10
31.	--	--	--	--	--	OCH_3	10
32.	--	--	--	--	--	$O(CH_2)_2NCH_3$	10
33.	--	--	--	--	--	SCH_3	10
34.	--	--	--	--	--	$SOCH_3$	10
35.	--	--	--	--	--	SO_2CH_3	10
36.	--	--	--	--	--	N_3	10
37.	--	--	--	--	--	NH_2	10
38.	--	--	--	--	--	$N(CH_3)_2$	10
39.	--	--	--	--	--	$CO_2C_2H_5$	10
40.	--	--	--	--	--	COOH	10
41.	--	--	--	--	--	CH_2OH	10
42.	--	--	--	--	--	CH_2OCOCH_3	10
43.	--	--	--	--	--	$CH_2OCOC_2H_5$	10

Comp. No.	R_1	R_2	R_3	R_4	Ar	X	Ref
44.	--	--	--	--	--	$CH_2OCO(CH_2)_4CH_3$	10
45.	--	--	--	--	--	$CH_2OCO(CH_2)_8CH_3$	10
46.	--	--	--	--	--	$CH_2OCO(CH_2)_{16}CH_3$	10
47.	--	--	--	--	--	$CH_2OCOCH_2N(CH_3)_2$	10
48.	--	--	--	--	--	$CH_2OCO(CH_2)_2N(CH_3)_2$	10
49.	--	--	--	--	--	$CH_2OCO(CH_2)_3N(CH_3)_2$	10
50.	--	--	--	--	--	CH_2OCH_3	10
51.	--	--	--	--	--	$CH_2O(CH_2)_2CH_3$	10
52.	--	--	--	--	--	$CH_2O(CH_2)_5CH_3$	10
53.	--	--	--	--	--	$CH_2O(CH_2)_2N(CH_3)_2$	10
54.	--	--	--	--	--	$CH_2O(CH_2)_3N(CH_3)_2$	10
55.	--	--	--	--	--	CH_2NH_2	10
56.	--	--	--	--	--	$CH_2N(CH_3)_2$	10
57.	--	--	--	--	--	$CH_2NHCO(CH_2)_4CH_3$	10
58.	--	--	--	--	--	--	11
59.	--	--	--	--		--	11
60.	--	--	--	--		--	11
61.	--	--	--	--		--	11
62.	--	--	--	--		--	11
63.	--	--	--	--	--	--	12
64.	phenyl	--	--	--	--	NH	12
65.	phenyl	--	--	--	--	NCH_3	12
66.	phenyl	--	--	--	--	NC_2H_5	12
67.	phenyl	--	--	--	--	$NCO_2C_2H_5$	12
68.	phenyl	--	--	--	--	CH_2	12
69.	phenyl	--	--	--	--	N(*p*-C_6H_4F)	12
70.	CH_3	--	--	--	--	N(*p*-C_6H_4F)	12
71.	phenyl	--	--	--	--	N(2-pyrimidyl)	12
72.	CH_3	--	--	--	--	N(2-pyrimidyl)	12
73.	CH_3	--	--	--	--	N(*p*-$C_6H_4OCH_3$)	12

TABLE-2
DATASET USED FOR PHARMACOPHORE-BASED 3D-QSAR ANALYSIS SHOWING THE ACTIVE AND INACTIVE PHARM SET WITH THEIR EXPERIMENTAL *Vs.* PREDICTED ACTIVITIES (pIC_{50}) USING THE BEST PHARMACOPHORE HYPOTHESIS **AADRR.22**

Ligand No.	Pharm Set	QSAR Set	Fitness	Exp. Activity	Predicted Activity	Ligand No.	Pharm Set	QSAR Set	Fitness	Exp. Activity	Predicted Activity
1	Inactive	Training	1.13	7	7.01	**38**	Active	Training	2.9	10.409	10.31
2	Inactive	Test	1.12	7	7.02	**39**		Training	2.44	8.854	8.39
3	Inactive	Training	2.05	7.523	7.62	**40**	Inactive	Training	2.49	8	8.53
4		Training	2.09	8.398	8.11	**41**	Active	Training	2.91	10.208	10.44
5		Training	2.08	8.398	8.41	**42**	Active	Test	2.88	10.585	9.12
6	Inactive	Training	2.09	7.222	7.32	**43**	Active	Test	2.86	9.886	10.04
7	Inactive	Training	2.08	7.155	7.11	**44**	Active	Training	2.76	9.796	9.51
8	Active	Training	1.34	9	9.17	**45**	Active	Training	2.68	10	10.22
9	Inactive	Test	1.65	7.301	8.23	**46**	Inactive	Test	2.53	8	9.74
10	Active	Training	2.09	9.097	8.39	**47**	Active	Training	2.8	10.585	10.66
11		Training	2.08	8.699	8.49	**48**	Active	Test	2.77	10.721	9.87
12		Test	2.11	8.155	8.25	**49**	Active	Training	2.74	10.538	10.83
13		Test	2.1	8.301	8.31	**50**	Active	Training	2.96	10.337	9.95
14	Inactive	Training	2.09	7.301	7.66	**51**		Training	2.88	8.149	8.53
15	Inactive	Test	2.04	7.301	8.19	**52**		Test	2.78	8.921	9.75
16	Inactive	Training	2.03	7.301	7.35	**53**	Inactive	Training	2.82	8	7.8
17	Inactive	Training	1.88	7.824	7.69	**54**	Inactive	Training	2.79	8	7.69
18		Training	2.04	8.155	7.96	**55**	Active	Training	3	10.585	10.4
19		Training	1.71	8.097	7.92	**56**	Active	Training	2.92	10.143	9.63
20	Inactive	Training	1.32	7.301	7.54	**57**	Active	Training	2.77	9.469	9.78
21	Inactive	Test	0.99	7.301	7.46	**58**		Training	1.61	8.26	8.26
22	Inactive	Training	1	7.301	7.44	**59**		Training	1.5	8.208	8.31

Ligand No.	Pharm Set	QSAR Set	Fitness	Exp. Activity	Predicted Activity	Ligand No.	Pharm Set	QSAR Set	Fitness	Exp. Activity	Predicted Activity
23	Active	Test	2.9	10.398	9.59	**60**		Training	1.51	8.569	8.8
24	Active	Training	2.49	10.886	10.22	**61**		Training	1.6	8.276	8.44
25	Active	Training	3	9.886	9.85	**62**		Training	1.5	8.553	8.62
26		Training	2.91	8.056	8.82	**63**	Active	Training	0.96	9.509	9.26
27	Inactive	Test	2.84	8	9.88	**64**		Test	2.08	8.228	8.25
28	Active	Training	2.88	9.678	10.17	**65**	Inactive	Training	2.06	7.65	7.75
29	Active	Test	2.49	9.796	9.94	**66**	Inactive	Training	0.97	7.246	7.27
30	Inactive	Training	2.49	8	9.39	**67**	Inactive	Training	2.02	7.868	7.95
31	Active	Test	2.5	10	9.53	**68**	Inactive	Test	2.08	7.666	8.37
32		Training	2.79	8.854	8.87	**69**		Training	2	8.561	8.94
33	Active	Training	2.88	9.721	9.66	**70**		Training	1.7	8.199	8.33
34		Test	2.49	8.602	8.61	**71**		Test	2.01	8.058	8.12
35	Inactive	Training	2.85	8	7.77	**72**	Inactive	Training	1.71	7.238	7.04
36	Active	Test	2.89	9.337	9.3	**73**		Test	1.69	8.47	8.02
37	Active	Training	2.49	11.114	9.61						

Active: pIC50 > 9; Inactive: pIC50 < 8

TABLE 3
SUMMARY OF QUANTITATIVE STRUCTURE-ACTIVITY RELATIONSHIP (QSAR) RESULTS FOR FIVE BEST COMMONPHARMACOPHORE HYPOTHESIS (CPHs)

ID	# Factors	SD	R^2	F	P	Stability	RMSE	Q^2	Pearson-R
AAHRR.42	1	0.8161	0.4877	45.7	1.721e-008	0.9242	0.7977	0.509	0.7326
	2	0.6931	0.6382	41.4	4.216e-011	0.8177	0.7191	0.601	0.791
	3	0.5843	0.7483	45.6	7.951e-014	0.5614	0.7249	0.5945	0.7852
	4	0.4926	0.825	53.0	1.823e-016	0.4613	0.7127	0.6081	0.7906
	5	0.4119	0.8803	64.7	3.648e-019	0.3568	0.7161	0.6043	0.7856
AADRR. 784	1	0.8467	0.4673	43.9	2.341e-008	0.9494	0.7949	0.5125	0.7204
	2	0.7121	0.6307	41.8	2.511e-011	0.6899	0.7916	0.5165	0.7236
	3	0.6285	0.7182	40.8	3.026e-013	0.5221	0.7703	0.5422	0.744
	4	0.5018	0.8241	55.1	3.725e-017	0.4079	0.7968	0.5101	0.7335
	5	0.4294	0.8739	63.8	1.524e-019	0.2891	0.8003	0.5058	0.728
ADHHR.876	1	0.8492	0.4641	43.3	2.717e-008	0.9714	0.7476	0.5688	0.7687
	2	0.7117	0.6311	41.9	2.454e-011	0.8582	0.7116	0.6093	0.7874
	3	0.6075	0.7367	44.8	5.967e-014	0.5954	0.6897	0.6329	0.8055
	4	0.5523	0.7869	43.4	3.243e-015	0.4668	0.7156	0.6049	0.7889
	5	0.4556	0.8581	55.6	2.261e-018	0.3203	0.6964	0.6258	0.8005
ADHHR.22	**1**	**0.8623**	**0.4474**	**40.5**	**5.969e-008**	**0.9819**	**0.7992**	**0.5072**	**0.7183**
	2	**0.6964**	**0.6468**	**44.9**	**8.435e-012**	**0.6319**	**0.6164**	**0.7068**	**0.848**
	3	**0.6205**	**0.7253**	**42.2**	**1.641e-013**	**0.6096**	**0.5201**	**0.7913**	**0.9108**
	4	**0.4934**	**0.8299**	**57.3**	**1.709e-017**	**0.4894**	**0.5521**	**0.7648**	**0.8779**
	5	**0.4396**	**0.8679**	**60.4**	**4.439e-019**	**0.3705**	**0.5038**	**0.8042**	**0.9044**
ADHHR.660	1	0.8772	0.4282	37.4	1.428e-007	0.975	0.7907	0.5176	0.7244
	2	0.7259	0.6163	39.3	6.431e-011	0.7666	0.6196	0.7038	0.8573
	3	0.63 0.55	0.7168	40.5	3.379e-013	0.6142	0.5663	0.7526	0.8871
	4	0.4742	0.7887	43.9	2.672e-015	0.5166	0.5497	0.7668	0.8822
	5		0.8463	50.7	1.393e-017	0.4205	0.5619	0.7563	0.8756

SD: Standard deviation of the regression, R^2:value of R^2 for the regression, F: variance ratio, ignificance level of variance ratio, RMSE: root-mean-square error, Q^2: value of Q^2 for the predicted activities, Pearson-R: Correlation between the predicted and observed activity for the test set

arrangements in 3D space. In the present study, 3D-pharmacophore hypothesis has been developed by employing a dataset of 73 molecules (Table-1) to understand the structural requirements for effective binding with the aromatase to design a series of new potent and selective human aromatase inhibitors.

Pharmacophore models containing five sites were generated using a terminal box size of 1 Å with 25 highly active molecules, belonging to reported aromatase inhibitors (Table-1), selected using a tree based partition algorithm. Prior to find the common pharmacophore hypothesis, the dataset was divided into active and inactive sets (Table-2). Pharmacophore models containing the 3, 4 and 6 featured common pharmacophore hypothesis were rejected and more than 6 sites could not be generated, as they were unable to define the complete binding space of the selected molecules.

In the present study, the common pharmacophore hypothesis that survived the scoring process were used to build an pharmacophore-based QSAR model and then evaluated and validated by correlating the observed and estimated activities by randomly selecting 30 % molecules as test sets (n = 21) using partial least square analysis. The PHASE descriptors served as independent variables and activity values as dependent variables in deducing 3D-QSAR models by partial least square regression analysis method. PHASE QSAR models do not use internal cross validation techniques but rather use distinct training and test sets. The predictive ability of each

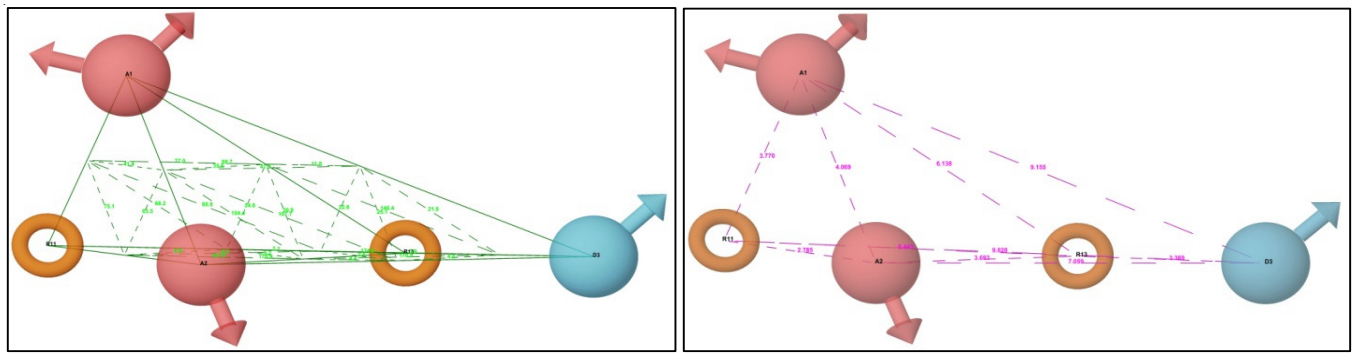


Fig. 1. Geometry of common pharmacophore hypothesis (CPH) **AADRR.22** (a)showing angle between various features (b) showing distances (Å) between various features.Pink spheres with vectors represents H-bond acceptor feature, orange torus represents aromatic ring feature, Blue spheres represents H-bond donor feature

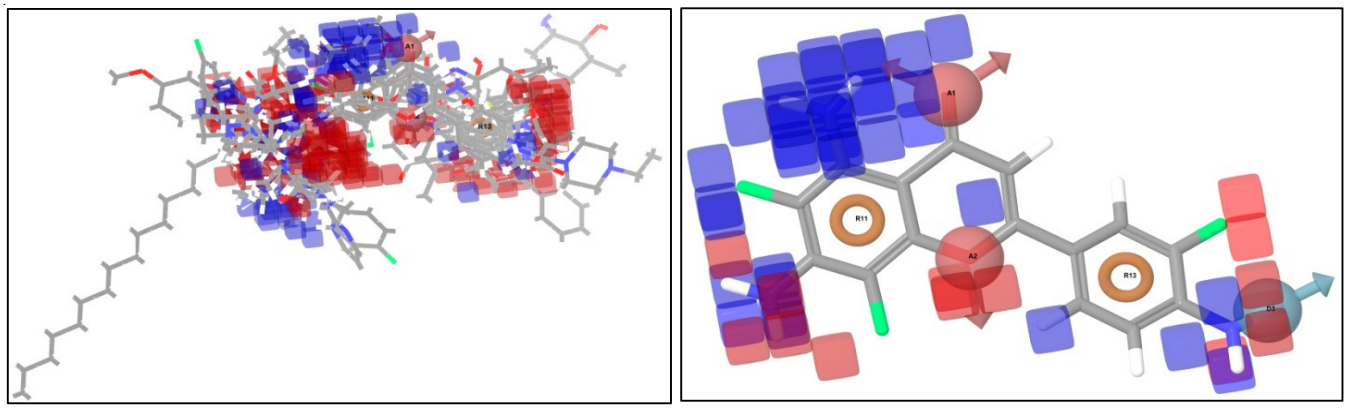


Fig. 2. (a) Alignment of all the ligands on pharmacophore hypothesis **AADRR.22** (b) Pictorial representation of the contours generated using QSAR model for most active compound **37**. Blue cubes indicate favourable regions while red cubes indicate unfavourable regions for the activity

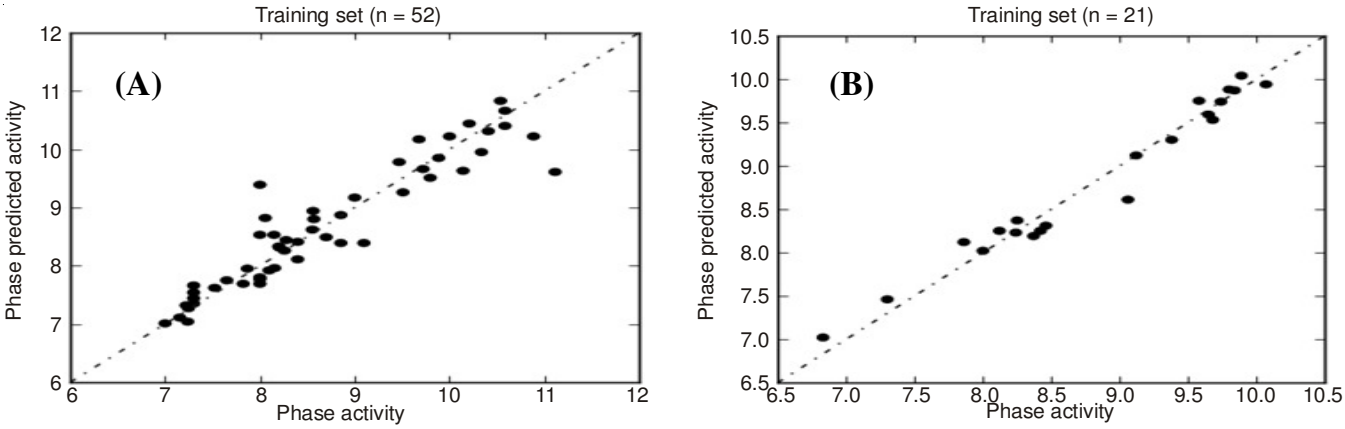


Fig. 3. (a) Scatter plot of predicted pIC_{50} (phase predicted activity) and experimental pIC_{50} (phase activity) of training set (n = 52) (b) scatter plot of predicted pIC_{50} (phase predicted activity) and experimental pIC_{50} (phase activity) of test set (n = 21)

analysis was determined from a test set of 21 molecules and these compounds were not included in the training set. The test set molecules were aligned and their activities were predicted by each partial least square analysis.

The common pharmacophore hypothesis with the best predictive value and significant statistical data were chosen for alignment of molecules. Total five hypotheses were found to be statistically very significant (Table-3). Finally the hypotheses namely **(AADRR.22)** (survival score = 3.774), selected (Fig. 1, Table-3) for molecular alignment. Partial least square analyses were conducted using five factors with a grid spacing of 1 Å.

The best pharmacophore model **(AADRR.22)** showed R^2 = 0.8679. The goodness of the model was validated by Q^2 (0.8042) and Pearson-R value (0.9044) for test set (Table-3). Other pharmacophore models with good statistics (Table-3) were not able to map diverse chemotypes having aromatase inhibitory activity. Hence, the hypothesis **(AADRR.22)** with two hydrogen bond acceptors (A), one hydrogen bond donor (D) and two aromatic ring (R) as pharmacophoric features was retained for further studies. Figure 2a shows the alignment of the active and inactive molecules under study along with **(AADRR.22)**. The distances and angles between the pharmacophoric features are depicted in Figs.1 and 2 respectively.

Analysis of the training set molecules aligned to the best pharmacophore hypothesis **(AADRR.22)** revealed two acceptors A1 & A2 mapping one on carbonyl oxygen atom (C=O) at 4th

position and other on cyclic oxygen atom (-O-) at 1st position of flavanone moiety, one hydrogen donor feature D3 mapping on the 4'-NH_2 group at the phenyl ring attached to 2nd position of flavanone moiety and two aromatic ring features R11 & R13 mapping on the fused benzene ring and phenyl ring attached to 2nd position of flavanone moiety (Fig. 2) with reference to the most active ligand **37**. Interestingly the alignment of all inactive molecules to the best pharmacophore hypothesis **(AADRR.22)** showed that they were unable to occupy the acceptor A1, A2 features and hydrogen bond donor D3 feature which are depicted to be the most crucial pharmacophoric features for showing the activity of the molecules.

The scatter plot of predicted *vs.* actual pIC_{50} for training and test set has been reported in Figs. 3a & 3b. It shows that all values were plotted around the best fit line indicating the significance of predicted model. The QSAR results can best be visualized using 3D plots of crucial pharmacophore regions (Fig. 2). The cubes (region) that represent the model are coloured according to the sign of their coefficient values. Blue cubes refer to region as positive coefficient, which is important for increase in activity, whereas red cubes refer to negative coefficients indicating the same ligand feature substitution with decrease in activity. The 3D plot of the most active molecule **37**, shows that optimum hydrophilic substitution at 5th position of flavanone moiety may enhance the activity. Interestingly, both the blue and red cubes surrounding the NH_2 group at 7th position of flavanone moiety indicate that presence of hydrophilic moiety like NH_2 group or availability of H-atom at this position is essential for retaining the activity, whereas substitution with other groups may be unfavourable and reduce the activity. Similarly, presence of NH_2 group at the 4'-position of the phenyl ring attached to 2nd position of flavonoid moiety is also essential as evident from the blue cubes (Fig. 3b). The above mentioned fact is also evident from another active molecule 23 of the dataset. Similarly, in case of pyrrolo[3,2-*f*]quinolin-9-one derivatives, the affinity of compound **10** is decreased as compared to the most active compound (**37**) due to the absence of NH_2 group at the 2nd position and 4'-position of 7-phenyl ring. The affinity was found to be further decreased in pyrrolo[3,2-*f*]quinolin-9-ones with substitution at the positions other than 4'-position of phenyl ring which falls into the unfavorable region (red cubes) as it was observed from compounds **11-15**. Further, lack of phenyl ring at the 7th position or with substitutions other than NH_2 group at the 2nd position resulted in decrease in affinity as seen in case of compounds **20-22**. In case of pyrroloquinolin-4-ones (compounds **2-9**), compounds without phenyl substituents at 2nd position (compounds **2, 3**) were found to have very less affinity as compared to those with phenyl substituents (compounds **4, 5**). Further, any substitution other than NH_2 group at 4'-position of phenyl ring also causes reduction in affinity. Pharmacophore based 3D-QSAR studies indicate that any molecule which satisfy the above pharmacophoric feature **(AADRR.22)** may have increase in binding affinity towards aromatase enzyme. Hence, based on this developed best pharmacophore hypothesis **(AADRR.22)**, novel potent aromatase inhibitors can be designed.

Conclusion

In the present study, a ligand-based 3D-QSAR pharmacophore model have been developed by considering 73 structurally diverse classes of heterocyclic compounds (IC_{50} = 0.0077-100 μM) which were found to have potent cytotoxic activity against MCF-7 cell lines (breast cancer cell lines) to understand the physicochemical requirements for effective binding with the aromatase enzyme to design a series of new potent and selective aromatase inhibitors (AIs). The 3D-QSAR pharmacophore hypotheses **(AADRR.22)** (survival score = 3.774) with two acceptors, one donor and two aromatic rings was found to be statistically most significant (SD = 0.4396, R^2 = 0.8679, F = 60.4 RMSE = 0.5038, Q^2 = 0.8042, Pearson-R = 0.9044) and could successfully identify known aromatase inhibitors and differentiate between active and inactive inhibitors. The mapping study of pharmacophore demonstrated that flavone derivatives could be the main scaffold on which substitutions at strategic positions may lead to increase the activity. The potential of this model can be effectively used to design new potent aromatase inhibitors for breast cancer treatment.

ACKNOWLEDGEMENTS

One of the authors D.P.K are thankful to Pharmaceutical Chemistry Division, School of Pharmacy, IMU, for providing the facilities to carry out the Cheminformatics (Docking) studies. The authors are also thankful to Mr. Raghurangaswamy, Executive Director, Schrodinger, India for providing the evaluation license (14 Jun-16 July, 2013) to carry out the molecular modeling studies.

REFERENCES

1. B.L. Narayana, D. Pran Kishore, C. Balakumar, K.V. Rao, R. Kaur, A.R. Rao, J.N. Murthy and M. Ravikumar, *Chem. Biol. Drug Des*., **79**, 674 (2012).
2. J. Narashimamurthy, A.R. Rao and G.N. Sastry, *Curr. Med. Chem. Anti-cancer Agents*, **4**, 523 (2004).
3. E.A. Ariazi, J.L. Ariazi, F. Cordera and V.C. Jordan, *Curr. Top. Med. Chem*., **6**, 181 (2006).
4. R.W. Brueggemeier, J.C. Hackett and E.S. Diaz-Cruz, *Endocr. Rev*., **26**, 331 (2005).
5. P. Roy and K. Roy, *J. Mol. Model*., **16**, 1597 (2010).
6. P. Roy and K. Roy, *J. Pharm. Pharmacol*., **62**, 1717 (2010).
7. Y. Dai, Q. Wang, X. Zhang, S. Jia, H. Zheng, D. Feng and P. Yu, *Eur. J. Med. Chem*., **45**, 5612 (2010).
8. M.G. Ferlin, G. Chiarelotto, V. Gasparotto, L. Dalla Via, V. Pezzi, L. Barzon, G. Palù and I. Castagliuolo, *J. Med. Chem*., **48**, 3417 (2005).
9. V. Gasparotto, I. Castagliuolo, G. Chiarelotto, V. Pezzi, D. Montanaro, P. Brun, G. Palù, G. Viola and M.G. Ferlin, *J. Med. Chem*., **49**, 1910 (2006).
10. T. Akama, H. Ishida, U. Kimura, K. Gomi and H. Saito, *J. Med. Chem*., **41**, 2056 (1998).
11. P. Dandawate, E. Khan, S. Padhye, H. Gaba, S. Sinha, J. Deshpande, K. Venkateswara Swamy, M. Khetmalas, A. Ahmad and F.H. Sarkar, *Bioorg. Med. Chem. Lett*., **22**, 3104 (2012).
12. M.N. Abu-Aisheh, M.S. Mustafa, M.M. El-Abadelah, R.G. Naffa, S.I. Ismail, M.A. Zihlif, M.O. Taha and M.S. Mubarak, *Eur. J. Med. Chem*., **54**, 65 (2012).
13. Maestro version 9.4 and Glide v5.9, Schrodinger, LLC, New York, 2013-1.
14. Phase Version 3.5. Schrodinger, LLC, New York, 2013-1.