

Fragment Based HQSAR Modeling and Docking Analysis of Conformationally Rigid 3-azabicyclo [3.1.0] Hexane Derivatives to Design Selective DPP-4 Inhibitors

Manjunath Ghate* and Shailesh V. Jain

Institute of Pharmacy, Nirma University, S. G. Highway, Ahmedabad 382481, Gujarat, India

Abstract: Development of potent, selective and orally bioactive dipeptidyl Peptidase IV inhibitors as antihyperglycemic agents is challenging task due to potential side effects are associated with them. It may result from other prolyldipeptidases of DPP-4 include DPP-2, DPP-8 and DPP-9. To resolve the selectivity issue in different DPP enzymes hologram quantitative structure-activity relationship studies were carried out on a series of potent and selective DPP-4 ligands. To measure selectivity between two kinds of enzyme selectivity data of DPP-4 over DPP-2, DPP-8 and DPP-9 were calculated and best HQSAR models were generated with significant correlation coefficients. The statistical results of the three models showed the best prediction and fitness for the selectivity activities. Docking studies were carried out on conformationally rigid 3-azabicyclo [3.1.0] hexane derivatives which suggested the substitution pattern on P1 and P2 fragment. The finally QSAR model, along with the information obtained from contribution maps and docking studies should be useful for the design of novel DPP-4 ligands having improved selectivity without side effects.

Keywords: HQSAR, 3-azabicyclo [3.1.0] hexane derivatives, Fragment based design, Selectivity, DPP-4 inhibitors, Docking.

INTRODUCTION

Type-2 diabetes is becoming a global epidemic, with the incidence and prevalence of the condition as well as related risk factors, such as obesity, continuously rising worldwide. This metabolic disease is a growing public health problem, affecting approximately 246 million people worldwide in 2007, and this number is projected to be 380 million by 2025 [1]. Within the past few years, increasing comprehension of the cellular and biochemical defects underlying insulin resistance and T2DM has prepared the foundations on which to develop new therapeutic approaches that enhance insulin secretion in a sustained glucose-dependent manner in patients with type 2 diabetes [2]. Although a number of therapies are available for this condition, recent efforts have been focusing on dipeptidyl peptidase 4 (DPP-4) inhibitors as a new class of therapeutic agents for type 2 diabetes [3]. Indeed, a number of clinical studies have already confirmed the efficacy and good tolerance of DPP-4 inhibitors.

The incretin hormones glucagon like peptide-1 (GLP-1) and glucose dependent insulinotropic polypeptide (GIP) play an important role in glucose homeostasis with effects on the pancreas, gastrointestinal tract, muscle tissue, and brain. GLP-1 enhances glucose-stimulated insulin secretion from the β -cells of the pancreas, promotes insulin biosynthesis, and inhibits postprandial glucagon secretion [4]. DPP-4 is a serine protease that cleaves a dipeptide from the N-terminus of the active form of GLP-1, GIP, neuropeptides, and chemokines, and renders them inactive [5]. This discovery has led to the development of DPP-4 inhibitors to increase

the half-life of circulating incretin hormones and normalize glucose homeostasis [6]. DPP-4 has now become a validated target with several small molecule inhibitors in late stage clinical trials for the treatment of type 2 diabetes. A number of catalytically active DPPs distinct from DPP-4 (DPP II, FAP, DPP-8, and DPP-9) have been described that is associated with side-effect and toxicity [7]. Although, a number of DPP-4 inhibitors have been accounted, all have limitations relating to potency, stability, or toxicity. Accordingly, a great need exists for novel DPP-4 inhibitors which are useful in treating conditions mediated by DPP-4 inhibition and which do not suffer from the above mentioned limitations [8]. The development of selective DPP-4 inhibitors is a big task due to another member of dpp family like DPP-2, DPP-8, DPP-9, etc. which may produce side effect as severe toxic reaction, alopecia, thrombocytopenia, anemia and increased mortality [9]. To discover potent and selective and safer drugs in a shorter time frame and with reduced cost it requires using an innovative approach for designing novel inhibitors. The utilization of CADD solutions and mathematical models like QSAR can help with future development of more selective DPP-4 agents.

Improved ligand based design methods that are based on such traditional QSAR methods e. g. fragment-based method HQSAR (Hologram QSAR) is effective to solve untreated issues [10]. Holographic QSAR (HQHQSAR) is a newly developed QSAR technique, which relates the biological activity to structural molecular composition in terms of patterns of sub-structural fragments. It eliminates the need for generation of 3D structure, putative binding conformations, and molecular alignment [11]. With the combined application of molecular hologram and subsequent partial least squares (PLS) regression analysis, highly predictive QSARs are developed and validated with cross-validation procedure. Leave-one-out cross-validation method is used for both fit-

*Address correspondence to this author at the Institute of Pharmacy, Nirma University, S. G. Highway, Ahmedabad 382481, Gujarat, India;
Tel: +91 - 2717 - 241900 to 04; Fax: +91 - 2717 - 241916;
E-mail: manjunath.ghate@nirmauni.ac.in

ting and assessing the model [12]. No 3D molecular structure and molecular alignment are needed for the generation of hologram. With partial least squares (PLS) regression analysis, the problem of co-linearity among parameters is avoided. In addition the molecular descriptors can be created automatically and quickly and avoid the selection and calculation or measurement of physicochemical descriptors required by traditional QSAR. Thus, it provides promising screening tools for large scale of dataset [13]. The results are viewed as color-coded regions of the individual molecules, the color-coding denoting regions of relatively high and low contributions to activities [14]. Applying the quantitative HQSAR model, predict the activity of a different (but structurally related) molecule and search databases of compounds similar to those in the training set. Docking Studies provide guidelines for the drug design and focus on the mechanisms of drug-receptor interactions, which further aid to the rationalization of drug development [15]. In the present work, HQSAR technique has been applied to generate molecular representation and derive QSAR model, aiming to develop robust, highly predictive QSAR models for design highly selective DPP-4 inhibitors by the measure of selectivity between the two kinds of enzymes. Docking studies were carried out for selective 3-azabicyclo [3.1.0] hexane derivatives against an anti DPP-4 target to explore the binding mechanism.

MATERIALS AND METHODS

Data Selection

A series of conformationally rigid 3-azabicyclo [3.1.0] hexane derivatives as selective DPP-4 are taken from literature (Table 1) [16]. To measure of selectivity between two kinds of enzyme selectivity data of DPP-4 over DPP-2, DPP-8 and DPP-9 were calculated by their respective biological activity (IC_{50}) using given formula:

$$p(IC_{50_{DPP4}}/IC_{50_{DPP2}}) = -\log (IC_{50_{DPP4}}/IC_{50_{DPP2}}) \dots (\text{HQSAR A})$$

$$p(IC_{50_{DPP4}}/IC_{50_{DPP8}}) = -\log (IC_{50_{DPP4}}/IC_{50_{DPP8}}) \dots (\text{HQSAR B})$$

$$p(IC_{50_{DPP4}}/IC_{50_{DPP9}}) = -\log (IC_{50_{DPP4}}/IC_{50_{DPP9}}) \dots (\text{HQSAR C})$$

All the molecules were built using the SybylX 1.2 molecular modeling package and were minimized by using the Tripos force field and Gasterger–Marsili charge with an energy gradient convergence criterion of 0.001 kcal/mol and a distance-dependent dielectric constant. The data sets were randomly partitioned into training and test set molecules by considering a range of molecules, so that both the training and test sets consist of high, medium and low activity molecules. The number of training and test set molecules consist as per Table 2, respectively.

Generation of Molecular Hologram

The novel molecular hologram representation designed by the HQSAR package of Sybyl X 1.2 software (Tripos Inc., St. Louis, USA). In HQSAR, each molecule in the database is divided into a set of unique overlapping structural fragments and sorted to form a molecular hologram, unlike other fragment-based fingerprinting methods, which encodes more information, such as branched and cyclic fragments and overlapping fragments as well as stereochemistry, and

maintains a count of the number of times about each fragment occurs. HQSAR model generation deals with the 2D structure directed fragment fingerprints. These molecular fingerprints are broken into strings at fixed intervals as specified by a hologram length (HL) parameter. The HL determines the number of bins in the hologram into which the fragments are hashed. The optimal HQSAR model was derived from screening through the 12 default HL values, which were a set of 12 prime numbers ranging from 53-401. The model development was performed using the following parameters: atom (A), bond (B), connection (C), chirality (Ch), hydrogen (H) and donor/acceptor (DA). The patterns of fragment counts from the training set inhibitors were then related to the experimental biological data using the PLS analysis.

QSAR Model Validation

All QSAR models were investigated using full cross-validated r^2 (q^2) PLS. Leave one-out (LOO) cross-validation has been applied to determine the number of principal components that yield optimally predictive models. External validation was performed with a test set compounds, which were not considered in QSAR model development. The predictive ability of the models is expressed by predictive r^2 values (r^2_{pred}), calculated as follows (equation 1):

$$r^2_{pred} = \frac{SD - PRESS}{SD} \quad (1)$$

SD is the sum of squared deviation between the biological activities of the test set molecule and the mean activity of the training set molecules and PRESS is the sum of squared deviations between the observed and the predicted activities of the test molecules [17].

Fisher's Weight

The Fisher weight is a measure of the distance between two categories; it is given by the difference of the mean values of each category, divided by the sum of the categories variances and can be interpreted as a normalized distance between the Classes [18]. The Fisher's weight is defined as eq. 2.

$$F = \frac{(\bar{x}_{p,1} - \bar{x}_{p,2})^2}{S_{p,1} + S_{p,2}} \quad (2)$$

Where $\bar{x}_{p,1}$, $\bar{x}_{p,2}$, denote the average values of descriptor p in class 1 and class 2 respectively, and $S_{p,1}$, $S_{p,2}$ denote the standard deviation of descriptor p in class 1 and class 2, respectively [19].

Docking Analysis

Degree of affection in between synthesized novel molecules and target in terms of structural and chemical complementation was explored by advanced scientific program "Glide 4.5" module of Schrödinger Molecular Modeling Interface [20]. Glide searches for favorable interactions between one or more ligand molecules and a receptor molecule using a grid based method. In the present study, X-ray crystal structure of DPP-4 (PDB: 1N1M) was taken from PDB (www.rcsb.org). Before docking the proteins were prepared

Table 1. Chemical Structure and their Biological Activities.

S.No.	R ¹	IC ₅₀ (nM)			
		DPP-4	DPP-2	DPP-8	DPP-9
1		165	18100	2110	421
2		153	2760	637	131
3		110	1200	3030	1660
4		38	10900	1675	343
5		95	1000	399	66
6		152	907	3020	807
7		59	1000	652	130
8		2540	n.d.	n.d.	n.d.
9		109	9000	500	771
10		229	9880	2080	640

Table 1. contd...

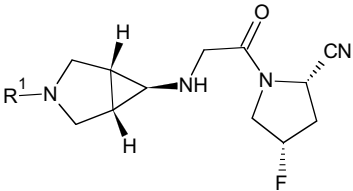
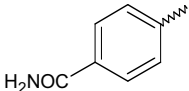
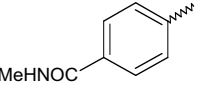
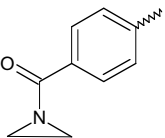
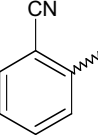
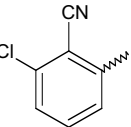
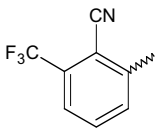
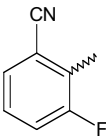
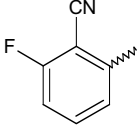
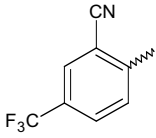
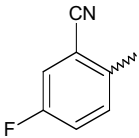
S.No.	R ¹	IC ₅₀ (nM)			
		DPP-4	DPP-2	DPP-8	DPP-9
					
11		95	21800	2130	3030
12		3430	n.d.	n.d.	n.d.
13		3740	n.d.	n.d.	n.d.
14		51	406	2400	729
15		31	96	10000	119
16		42	1300	6500	1910
17		63	1300	366	169
18		92	>10000	10000	1910
19		141	4330	28600	3680
20		422	>10000	3090	4850

Table 1. contd....

S.No.	R ¹	IC ₅₀ (nM)			
		DPP-4	DPP-2	DPP-8	DPP-9
21		630	3900	2050	1960
22		118	1800	3270	1200
23		1715	n.d.	n.d.	n.d.
24		84	16500	4030	100000
25		181	2200	11100	5420
26		146	11000	11499	448
27		186	226	7270	7810
28		186	5100	14900	547
29		221	5100	13100	5990
30		3670	n.d.	n.d.	n.d.

Table 1. contd...

S.No.	R ¹	IC ₅₀ (nM)			
		DPP-4	DPP-2	DPP-8	DPP-9
31		27	10800	837	n.d.
32		49	8750	2375	n.d.
33		85	13050	1667	n.d.
34		123	4000	35500	63200
35		147	11000	2180	350
36		78	6420	3400	525
37		91	1340	16100	7300
38		1300	n.d.	n.d.	n.d.

by using the protein preparation wizard, removing the water molecule and cofactors from the proteins, optimizing hydrogen bonding and deleting the ligand present in crystal structure. Solvent molecules were deleted and bond order for crystal ligand and protein were adjusted and minimized up to 0.30 Å RMSD. The ligands were built using Maestro v-8.0³ build panel and prepared by LigPrep-v2.1 [21] by using the OPLS-2005 [22] force field. Using extra precision (XP)

mode of Glide v-4.5 all molecules were docked into the active site of target molecules and final scoring is carried out in term of GlideScore multi-ligand scoring function.

RESULTS AND DISCUSSION

HQSAR is a technique that employs fragment fingerprints as predictive variables of biological activity or other

Table 2. Training Set and Test Set Distinction of Developed Models.

Model	Total Compounds	Activity Range (PIC ₅₀)	Training Set	Test Set
(A) p(IC ₅₀ _{DPP4} /IC ₅₀ _{DPP2})	32	9.08-11.60	24	8
(B) p(IC ₅₀ _{DPP4} /IC ₅₀ _{DPP8})	32	9.62-11.51	24	8
(C) p(IC ₅₀ _{DPP4} /IC ₅₀ _{DPP9})	29	8.84-12.08	22	7

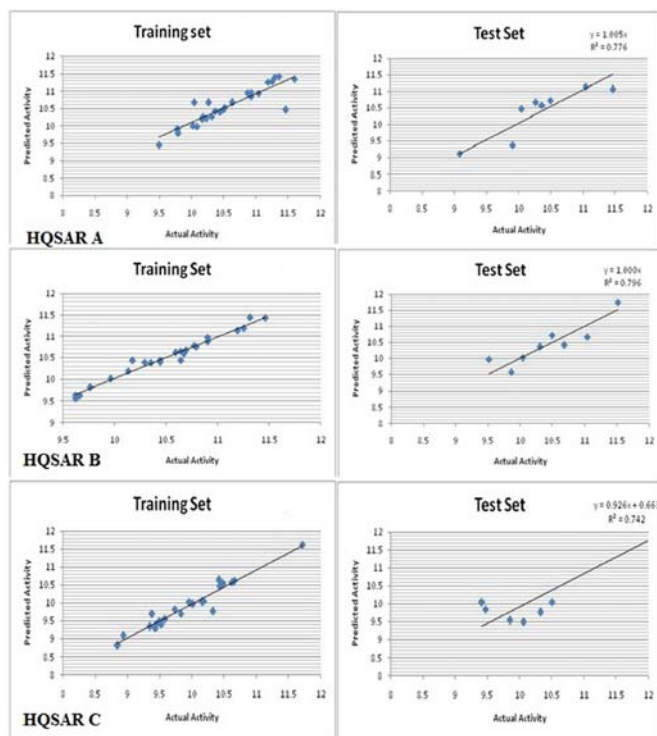


Fig. (1). Relationships between observed inhibition values and predicted values by HQSAR models. The scatterplot displays the predicted versus actual activities of the training set (A) and the test set (B) of compounds.

structural related data HQSAR model generation deals with the 2D structure directed fragment fingerprints. The model development was performed using the following parameters: atom (A), bond (B), connection (C), chirality (Ch), hydrogen (H) and donor/acceptor (DA). The validity of the model depends on statistical parameters such as r^2 , q^2 by LOO, predictive r^2_{pred} and standard error. The robustness of the model depend on the more challenging test set prediction reflected by its predictive r^2_{pred} value. The results of HQSAR analysis indicate significance differences in HQSAR model quality when fragment parameter and fragment length is varied. The optimized model was generated with different combination of fragment parameter and fragment length and validate by cross-validation for HQSAR A, HQSAR B and HQSAR C. HQSAR analyses were done using various combinations of fragment distinctions (A, B, C, Ch, H, DA) using fragment size default 5-8 and best model was selected based on cross-validated predictability (q^2). The HQSAR results represent the best predictive model of the system, for the given choice of hologram parameters. Applying the quantitative model thus obtained, the activity of a different molecule can be predicted.

The HQSAR model (A-C) together with the results of final PLS calculation between the inhibition activities and

fragment size based on the distinctions of substrate molecules was visually analyzed by means of atomic contribution plots. Following the calculation of atomic contributions to activity, the molecule is color coded to reflect the individual atomic contributions. The colors at the red end of the spectrum (red and orange) reflect poor (or negative) contributions, while colors at the green end (yellow, blue, and green) reflect favorable (positive) contributions. Atoms with intermediate contributions are colored white. Comparison of the predicted and experimental results of the HQSAR model and their deviations are listed in Table 2. For HQSAR models (A-C), the Obs.PI50 versus the predicted Pred.PI50 values are plotted in Fig. (1).

HQSAR A

The results shown in Table 4a for HQSAR A conforming that model A3 is best model having good non cross validated (r^2) = 0.978 and cross validated (q^2)=0.711 having fragment distinction Atom, bond, connection and hydrogen bond. Further taking fragment distinction Atom, bond, connection and hydrogen bond, fragment length are varied from 2-5 to 8-12. The cross-validated predictability (q^2) varied from 0.525 at a fragment length of 2-5 to 0.711 at a fragment length of 5-8; further increases in fragment size decrease the q^2 . Cross-

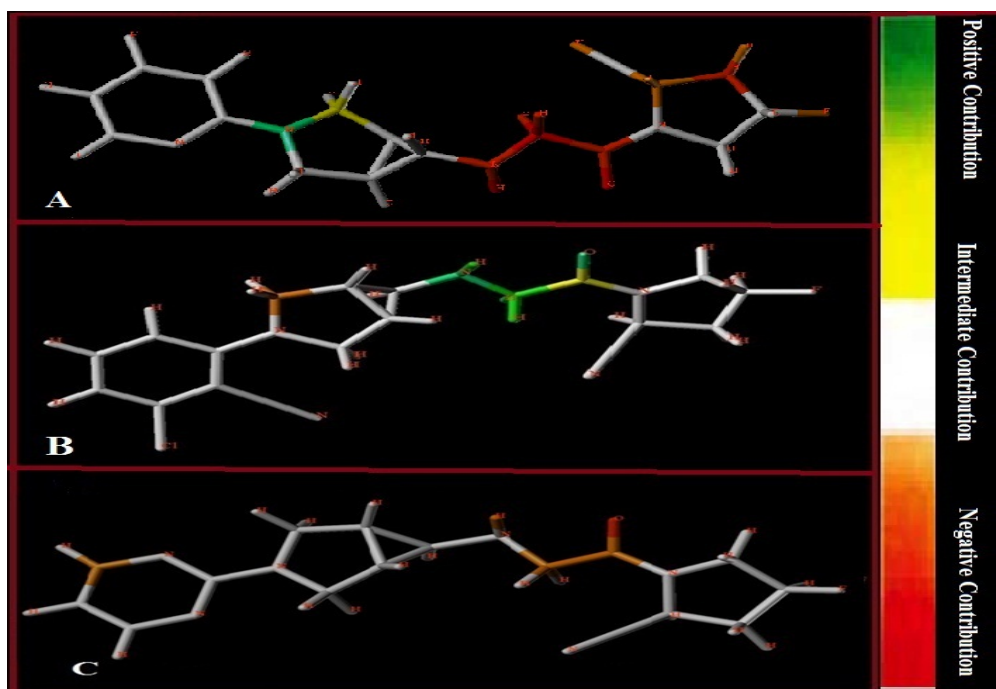


Fig (2). Individual atomic contributions of most active compound for the activity (A) compound 31 for HQSAR A model (B) compound 15 for HQSAR B Model(C) compound 24 for HQSAR C model.

validated and non cross-validated PLS analysis suggested that model A3 is the best HQSAR model (Table 4b). The results of the external validation for the best predictive model is $r^2_{\text{pred}} = 0.776$ and the graphic results for the experimental versus predicted for both training and test sets are shown in Fig. (1). The Atomic contribution to the activity of individual fragment is calculated using A3 model and denoted by color coding shown in Fig. (2). Fig (2A) showing atomic contribution map of highly selective molecule 31 in which red to brown color shows negative contribution, white color shows intermediate contribution and green color shows positive contribution towards selectivity of dpp-4 activity on dpp-2 activity. Contribution map shows that P1 fragment 4-fluoropyrrolidine-2-carbonitrile and peptide linkage in molecule contributing negatively as well as 3-azabicyclo [3.1.0]hexane ring and aromatic P2 fragment shown intermediate and positive contribution. Overall electronegative substitution at P1 fragment decreases the selectivity and P2 fragment increases the selectivity.

HQSAR B

The results shown in Table 5a for HQSAR B has shown that model B3 is best model having good non cross validated (r^2) = 0.973 and cross validated (q^2)=0.795 having fragment distinction Atom, bond, connection and hydrogen bond. Further taking fragment distinction Atom, bond, connection and hydrogen bond, fragment length are varied from 2-5 to 8-11. The cross-validated predictability (q^2) varied from 0.354 at a fragment length of 2-5 to 0.795 at a fragment length of 5-8; further increases in fragment size decrease the q^2 . Cross-validated and non cross-validated PLS analysis suggested that model B3 is the best HQSAR model with fragment size 5-8 and fragment distinction Atom, bond, connection and hydrogen bond (Table 5b). The results of the external valida-

tion for the best predictive model is $r^2_{\text{pred}} = 0.796$ and the graphic results for the experimental versus predicted for both training and test sets are shown in Fig. (1). The Atomic contribution to the activity of individual fragment is calculated using the B3 model and denoted by color coding shown in Fig. (2). Fig. (2B) showing atomic contribution map of highly selective molecule 15 in which red to brown color shows negative contribution, white color shows intermediate contribution and green color shows positive contribution towards selectivity of dpp-4 activity on dpp-8 activity. Contribution map shows that peptide linkage in molecule contributing positively as well as P1 fragment 4-fluoropyrrolidine-2-carbonitrile and aromatic P2 fragment shown intermediate contribution. 3-azabicyclo[3.1.0]hexane ring shown negative contribution. An overall linkage group of P1 and P2 fragment is important for selectivity of DPP-4 inhibitors.

HQSAR C

The results shown in Table (6a) for HQSAR B has shown that model C3 is best model having good non cross validated (r^2) = 0.968 and cross validated (q^2)=0.646 having fragment distinction Atom, bond, connection and hydrogen bond at 5-8 fragment length. Further taking fragment distinction Atom, bond, and hydrogen bond, fragment length are varied from 2-5 to 8-11. The cross-validated predictability (q^2) varied from 0.394 at a fragment length of 2-5 to 0.646 at a fragment length of 5-8; further increases in fragment size decrease the q^2 . Cross-validated and non cross-validated PLS analysis suggested that model C6 is the best HQSAR model with fragment size 5-8 and fragment distinction Atom, bond and hydrogen bond (Table 6b). The results of the external validation for the best predictive model is $r^2_{\text{pred}} = 0.742$ and the graphic results for the experimental versus predicted for

Table 3. Actual and Predicted Activity of Developed Models.

Comp.	HQSAR A p(IC50 _{DPP4} /IC50 _{DPP2})			HQSAR B p(IC50 _{DPP4} /IC50 _{DPP8})			HQSAR C p(IC50 _{DPP4} /IC50 _{DPP9})		
	Actual	Pred.	Residual	Actual	Pred.	Residual	Actual	Pred.	Residual
1	11.04	10.928	0.112	10.13	10.179	-0.049	9.41	10.05	-0.64
2	10.26	10.667	-0.407	9.62	9.548	0.072	8.93	9.107	-0.177
3	10.04	10.468	-0.428	10.44	10.391	0.049	10.18	10.056	0.124
4	11.46	11.061	0.399	10.64	10.64	0	9.96	10.028	-0.068
5	10.02	10.01	0.01	9.62	9.632	-0.012	8.84	8.815	0.025
6	9.78	9.919	-0.139	10.3	10.358	-0.058	9.73	9.822	-0.092
7	10.23	10.21	0.02	10.04	10.027	0.013	9.34	9.338	0.002
8	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
9	10.92	10.842	0.078	9.66	9.622	0.038	9.85	9.535	0.315
10	10.63	10.663	-0.033	9.96	10.005	-0.045	9.45	9.422	0.028
11	11.36	11.413	-0.053	10.35	10.372	-0.022	10.5	10.05	0.45
12	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
13	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
14	9.9	9.377	0.523	10.67	10.589	0.081	10.16	10.036	0.124
15	9.49	9.438	0.052	11.51	10.741	0.769	9.58	9.542	0.038
16	10.49	10.728	-0.238	11.19	11.128	0.062	10.66	10.629	0.031
17	10.31	10.263	0.047	9.76	9.812	-0.052	9.43	9.307	0.123
18	11.04	11.138	-0.098	11.04	10.68	0.36	10.32	9.759	0.561
19	10.49	10.477	0.013	11.31	11.432	-0.122	10.42	10.631	-0.211
20	10.37	10.427	-0.057	9.86	9.571	0.289	10.06	9.483	0.577
21	9.79	9.791	-0.001	9.51	9.972	-0.462	9.49	9.465	0.025
22	10.18	10.261	-0.081	10.44	10.426	0.014	10.01	9.978	0.032
23	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
24	11.29	11.394	-0.104	10.68	10.41	0.27	12.08	12.296	-0.216
25	10.08	9.989	0.091	10.79	10.744	0.046	10.48	10.545	-0.065
26	10.88	10.905	-0.025	10.9	10.965	-0.065	9.49	9.459	0.031
27	9.08	9.121	-0.041	10.59	10.614	-0.024	10.62	10.568	0.052
28	10.44	10.405	0.035	10.9	10.871	0.029	9.47	9.855	-0.385
29	10.36	10.578	-0.218	10.77	10.775	-0.005	10.43	10.479	-0.049
30	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
31	11.6	11.341	0.259	10.49	10.715	-0.225	n.d.	n.d.	n.d.
32	11.25	11.28	-0.03	10.69	10.667	0.023	n.d.	n.d.	n.d.
33	11.19	11.249	-0.059	10.29	10.381	-0.091	n.d.	n.d.	n.d.
34	10.51	10.505	0.005	11.46	11.422	0.038	11.71	11.616	0.094
35	10.87	10.951	-0.081	10.17	10.428	-0.258	9.38	9.703	-0.323

Table 3. contd...

Comp	HQSAR A p(IC50 _{DPP4} /IC50 _{DPP2})			HQSAR B p(IC50 _{DPP4} /IC50 _{DPP8})			HQSAR C p(IC50 _{DPP4} /IC50 _{DPP9})		
	Actual	Pred.	Residual	Actual	Pred.	Residual	Actual	Pred.	Residual
36	10.92	10.951	-0.031	10.64	10.428	0.212	9.83	9.703	0.127
37	10.17	10.191	-0.021	11.25	11.179	0.071	9.52	9.403	0.117
38	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.

*Grey color indicates test set compounds

Table 4a. The HQSAR Analyses of Model HQSAR A for Various Fragment Distinction on the Key Statistical Parameters Using Fragment Size Default 5-8.

Model	Fragment Distinction	R2	SEE	Q2	SEP	HL	N
A1	A/B	0.973	0.107	0.677	0.370	257	6
A2	A/B/C	0.923	0.170	0.639	0.370	53	4
A3	A/B/C/H	0.978	0.096	0.711	0.350	97	6
A4	A/B/C/H/Ch	0.980	0.093	0.672	0.373	97	6
A5	A/B/C/H/Ch/DA	0.958	0.134	0.495	0.463	307	6
A6	A/B/H	0.979	0.095	0.556	0.434	307	6
A7	A/B/C/Ch	0.921	0.173	0.632	0.374	53	4
A8	A/B/DA	0.948	0.148	0.535	0.444	53	6
A9	A/B/C/DA	0.957	0.135	0.560	0.432	257	6
A10	A/B/H/DA	0.944	0.154	0.485	0.467	83	6
A11	A/B/C/Ch/DA	0.955	0.138	0.557	0.434	257	6
A12	A/B/C/H/DA	0.946	0.152	0.519	0.452	257	6
A13	A/B/H/Ch/DA	0.975	0.104	0.476	0.472	307	6

Table 4b. Influence of Various Fragment Sizes on Key Statistical Parameters Using the Best Fragment Distinction (A, B, C and H) of Model HQSAR A.

Fragment Size	R2	SEE	Q2	SEP	HL	N
2-5	0.929	0.174	0.525	0.449	71	6
3-6	0.947	0.150	0.582	0.421	97	6
4-7	0.983	0.084	0.657	0.381	401	6
5-8	0.978	0.096	0.711	0.350	97	6
6-9	0.974	0.104	0.556	0.429	401	6
7-10	0.975	0.104	0.556	0.429	401	6
8-11	0.894	0.201	0.459	0.453	61	4

Table 5a. The HQSAR Analyses of Model HQSAR B for Various Fragment Distinction on the Key Statistical Parameters Using Fragment Size Default 5-8.

Model	Fragment Distinction	R2	SEE	Q2	SEP	HL	N
B1	A/B	0.959	0.126	0.58	0.403	199	6
B2	A/B/C	0.969	0.109	0.602	0.393	257	6
B3	A/B/C/H	0.973	0.102	0.795	0.283	83	6
B4	A/B/C/H/Ch	0.971	0.107	0.762	0.304	83	6
B5	A/B/C/H/Ch/DA	0.953	0.135	0.584	0.402	71	6
B6	A/B/H	0.941	0.148	0.650	0.359	59	5
B7	A/B/C/Ch	0.970	0.109	0.623	0.383	257	6
B8	A/B/DA	0.914	0.173	0.616	0.366	257	5
B9	A/B/C/DA	0.959	0.126	0.566	0.414	353	6
B10	A/B/H/DA	0.956	0.132	0.598	0.395	257	6
B11	A/B/C/Ch/DA	0.952	0.137	0.545	0.421	307	6
B12	A/B/C/H/DA	0.948	0.142	0.594	0.397	97	6
B13	A/B/H/Ch/DA	0.927	0.163	0.584	0.391	353	5

Table 5b. Influence of Various Fragment Sizes on Key Statistical Parameters using the Best Fragment Distinction (A, B, C and H) of Model HQSAR B.

Fragment Size	R2	SEE	Q2	SEP	HL	N
2-5	0.627	0.343	0.354	0.451	71	2
3-6	0.957	0.126	0.601	0.383	71	5
4-7	0.966	0.115	0.684	0.351	83	6
5-8	0.973	0.102	0.795	0.283	83	6
6-9	0.952	0.129	0.681	0.333	83	4
7-10	0.885	0.201	0.617	0.365	71	4
8-11	0.942	0.146	0.628	0.370	307	5

Table 6a. The HQSAR Analyses of Model HQSAR C for Various Fragment Distinction on the Key Statistical Parameters Using Fragment Size Default 5-8.

Model	Fragment Distinction	R2	SEE	Q2	SEP	HL	N
C1	A/B	0.90	0.232	0.201	0.654	199	4
C2	A/B/C	0.65	0.408	0.177	0.628	53	2
C3	A/B/C/H	0.707	0.358	0.239	0.620	53	3
C4	A/B/C/H/Ch	0.700	0.390	0.226	0.626	53	3
C5	A/B/C/H/Ch/DA	0.894	0.239	0.296	0.614	401	4
C6	A/B/H	0.968	0.145	0.646	0.480	53	7
C7	A/B/C/Ch	0.886	0.247	0.172	0.666	97	4
C8	A/B/DA	0.832	0.300	0.140	0.679	59	4

Table 6a. cont...

Model	Fragment Distinction	R2	SEE	Q2	SEP	HL	N
C9	A/B/C/DA	0.923	0.217	0.260	0.670	83	6
C10	A/B/H/DA	0.886	0.254	0.155	0.694	353	5
C11	A/B/C/Ch/DA	0.875	0.267	0.180	0.683	53	5
C12	A/B/C/H/DA	0.898	0.241	0.240	0.658	401	5
C13	A/B/H/Ch/DA	0.894	0.239	0.296	0.614	401	4

Table 6b. Influence of Various Fragment Sizes on Key Statistical Parameters Using the Best Fragment Distinction (A, B, and H) of Model HQSAR C.

Fragment size	R2	SEE	Q2	SEP	HL	N
2-5	0.725	0.373	0.393	0.554	61	3
3-6	0.937	0.196	0.450	0.578	59	6
4-7	0.903	0.235	0.423	0.573	56	5
5-8	0.968	0.145	0.646	0.480	53	7
6-9	0.949	0.176	0.583	0.503	53	6
7-10	0.910	0.223	0.440	0.583	53	6
8-11	0.730	0.370	0.299	0.595	59	3

Table 7. Docking Scores.

Ligand	GScore	LipophilicEvdW	PhobEn	HBond	Electro
1	-9.13	-1.64	-1.4	-2.54	-2.97
2	-9.4	-1.35	-1.5	-2.99	-3.11
3	-8.49	-1.36	-1.38	-2.34	-2.92
4	-9.89	-1.4	-1.51	-2.97	-3.16
5	-9.36	-1.53	-1.42	-2.84	-3.05
6	-9.62	-1.92	-1.4	-2.71	-3.06
7	-9.25	-1.58	-1.6	-2.71	-3.07
8	-9.28	-1.6	-1.45	-2.83	-3
9	-8.64	-1.4	-1.46	-2.46	-3.33
10	-8.98	-1.79	-1.43	-2.74	-2.92
11	-8.67	-1.11	-1.35	-2.7	-3.38
12	-9.55	-1.68	-1.38	-3.01	-3.1
13	-9.29	-1.87	-1.51	-2.92	-2.95
14	-8.83	-1.39	-1.39	-2.61	-2.92
15	-8.94	-1.49	-1.43	-2.58	-2.92
16	-8.76	-1.28	-1.52	-2.62	-3
17	-9.45	-1.63	-1.46	-2.81	-3.03
18	-9.29	-1.56	-1.43	-2.73	-3.09

Table 7. contd...

Ligand	GScore	LipophilicEvdW	PhobEn	HBond	Electro
19	-9.38	-1.82	-1.39	-2.45	-3.14
20	-8.93	-1.38	-1.46	-2.63	-2.94
21	-9.2	-1.51	-1.36	-2.7	-3.06
22	-8.89	-1.53	-1.4	-2.68	-3.05
23	-8.39	-1.54	-1.38	-2.23	-2.78
24	-8.93	-1.34	-1.4	-2.68	-3.29
25	-9.05	-1.39	-1.38	-2.68	-2.93
26	-9.14	-1.37	-1.41	-2.85	-3.34
27	-8.8	-1.45	-1.38	-2.66	-2.93
28	-8.98	-1.54	-1.38	-2.64	-2.91
29	-9.01	-1.42	-1.4	-2.83	-3.36
30	-9.04	-1.44	-1.45	-2.63	-3
31	-8.98	-1.26	-1.39	-2.78	-2.97
32	-9	-1.34	-1.4	-2.82	-3.3
33	-8.37	-1.72	-1.4	-1.76	-3.18
34	-9.05	-1.39	-1.38	-2.67	-2.94
35	-8.99	-1.32	-1.49	-2.69	-2.87
37	-8.51	-1.34	-1.32	-2.15	-2.89
38	-8.71	-1.41	-1.41	-2.5	-2.93
Vildagliptin	-8.51	-1.24	-1.3	-3.09	-3.14
NVP	-6.95	-1.23	-1.04	-2.32	-3.08
saxagliptin	-9.6	-1.93	-1.16	-3.1	-3.63
sitagliptin	-7.92	-1.36	-1.36	-1.38	-3.26

both training and test sets are shown in Fig. (1). The Atomic contribution to the activity of individual fragment is calculated using C6 model and denoted by color coding shown in Fig. (2). Fig. (2C) showing atomic contribution map of highly selective molecule 15 in which red to brown color shows negative contribution, white color shows intermediate contribution and green color shows positive contribution towards selectivity of dpp-4 activity on dpp-9 activity. Contribution map shows that peptide linkage in molecule contributing negatively as well as P1 fragment 4-fluoropyrrolidine-2-carbonitrile and 3-azabicyclo [3.1.0] he-xane ring shows intermediate contribution. Aromatic P2 fragment shows negative contribution. Overall, peptide linkage involved in selectivity of dpp-4 inhibitors over dpp-9 activity.

The HQSAR models are dependent not only on the molecular hologram length, but also on the information including molecular fragment and distinction as before. Therefore, based on the selected hologram length, it is possible to find the structure of the substrate molecule and distinctions showing high inhibitory activity by searching either minimizing

the lowest standard error or the highest predictability q^2 . Finally, a complete HQSAR analysis involves the investigation of important indications of the molecular fragments directly related to biological activity or responsible for the low biological potency of the compounds and to propose structural modifications. In this way, one can obtain contribution maps that indicate the individual contributions to the activity of each atom in a given molecule of the data set and to analysis the most relevant structural fragments incorporated.

Docking Analysis

To investigate the detailed intermolecular interactions between the ligand and the target protein, an automated docking program Glide was used. These studies helped to sort out the screened compounds. Three-dimensional structure information on the DPP-4 enzyme was taken from the PDB entry 1N1M having resolution of 2.5 Å. DPP-4 is a 766 amino acid transmembrane glycoprotein which belongs to the prolyl oligopeptidase family. It consists of three parts; a cytoplasmic tail, a transmembrane region and an extracellu-

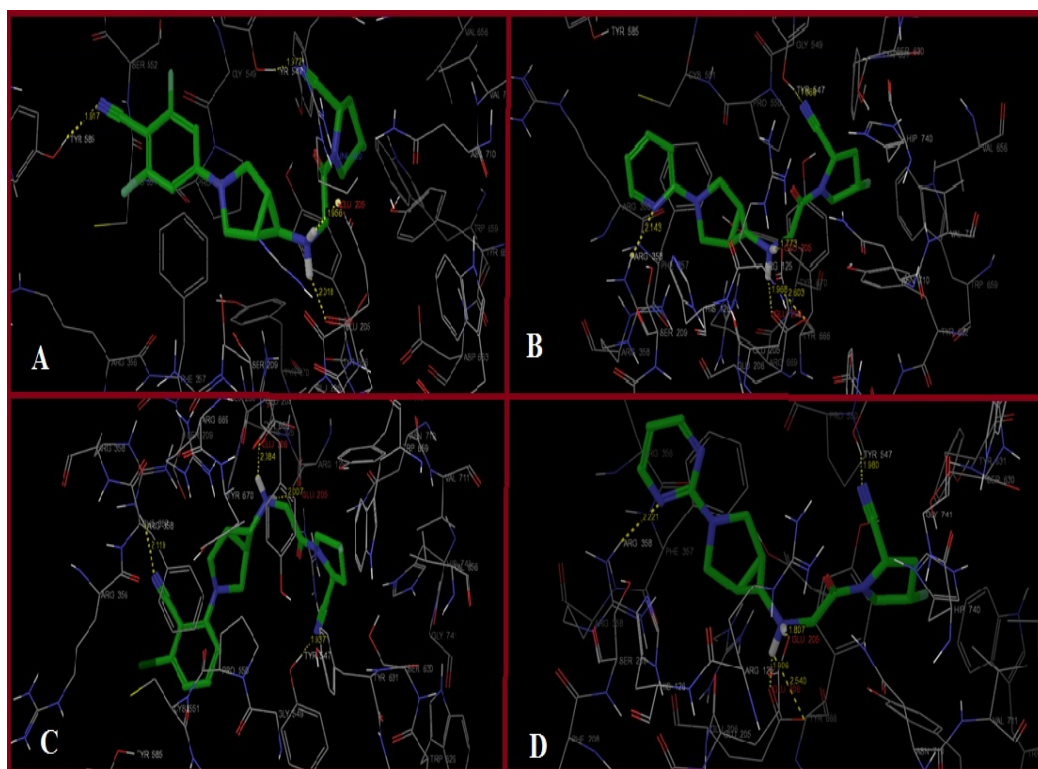


Fig. (3). Docking interaction of target enzyme dpp-4 (PDB: 1N1M) and series compounds (A) pose of highest docking score compound 4 (B) most active compound 31 for hqsar A model (C) most active compound 15 for hqsar B model (D) most active compound 15 for hqsar C model. Dotted yellow bond showing H-bond interactions with binding site residues.

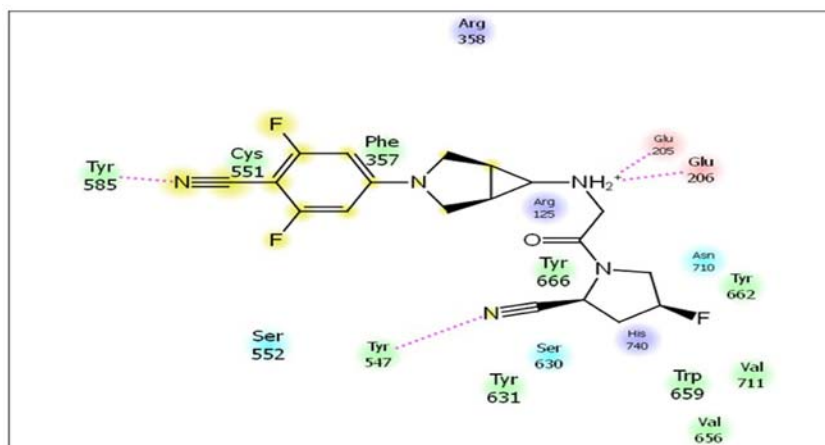


Fig. (4). 2D interaction diagram of highest docking score compound 4.

lar part. The extracellular part is divided into a catalytic domain and an eight-bladed α -propeller domain. The latter contributes to the inhibitor binding site. The catalytic domain shows α/β -hydrolase fold and contains the catalytic triad Ser630 - Asp708 - His740. The S1-pocket is very hydrophobic and is composed of the side chains: Tyr631, Val656, Trp662, Tyr666 and Val711. Existing X-ray structures show that there is not much difference in size and shape of the pocket that indicates that the S1-pocket has high specificity for Proline residue.

Protein preparation included the deletion of the ligand and the solvent molecules as well as the addition of hydrogen atoms and energy minimization. All conformationally rigid 3-azabicyclo[3.1.0]hexane derivatives were docked into

target using Glide extra precision docking module of Maestro 9.0 software. Evaluation is done with glide score (docking score) and single best pose is generated as the output for particular ligand. The entire compound showing good Glide score in the range of -9.89 to -8.37 as compared to establish drug for dpp-4 target. In the series of 3-azabicyclo [3.1.0]hexane derivatives compound 4 showing highest Gscore -9.89, lipophilic score -1.4, hydrophobic score -1.51, hydrogen bonding score -2.97 and electrostatic score -3.16 having good binding affinity to dpp-4 enzyme shown in Fig. (3a). Same binding pattern of most selective compound 31, 15 and 24 were shown in Fig. (3B-D).

Docking studies reveals the proline mimic P-1 site contains nitrile in the position of the scissile bond of the peptidic

substrate group forms reversible covalent bonds with the catalytically active Tyr 547 residue. 2 (S) -cyanopyrrolidine moiety has been found to be an integral part of many DPP-4 inhibitors and important for higher potency. The Hydrogen bonding network is formed between the protonated amino group and a negatively charged region of the protein surface, Glu205, Glu206 and Tyr662. 3-azabicyclo [3.1.0]hexane worked as a steric bulk and slowed intramolecular cyclization with increasing chemical stability. As P2 cavity have higher volume so bulky group were well tolerated at P2 site, different substituted aromatic and hetro aromatic show good lipophilic score. Elecronegetive group presents in ortho and pera position at P2 site form hydrogen bonding Arg 358 and Tyr 585 and well tolerated. In summary, structure-based design utilizes to build novel series DPP-4 inhibitors from a small fragment lead. SAR development led to the discovery of multiple compounds which are potent and selective while maintaining excellent physical properties and drug-like characteristics.

CONCLUSION

With the recent identification of several closely related proline-specific DPP enzymes, understanding the degree of selectivity required for the development of inhibitors with an optimal safety profile has become a key issue. In this study, a highly predictive HQSAR model for selective DPP-4 inhibitors was generated. The reliable model showed leave-one-out cross-validation q^2 and conventional r^2 values of 0.0711 and 0.978 for DPP-4 inhibitors, 0.79 and 0.973 for DPP-8 inhibitors, and 0.96 and 0.64 for DPP-9 inhibitors with respect to selectivity. The reliability of the HQSAR model was verified by the compounds in a test set with higher predictive power.

From the atomic contribution maps obtained using the HQSAR model, it was revealed that ortho or pera electro-withdrawing substituent's on the phenyl ring at S2 site contributed to the inhibition activity. According to this study, in order to obtain selective DPP-4 inhibitors compared to the isozymes, the interaction of the inhibitors with the S2 site and S1 site in DPP-4 should be carefully considered. To improve the potency as well as selectivity and stability it's required to optimize lead molecules by rational way, in such context ligand-based lead optimization is a powerful approach to the selection of lead compounds with potential for drug development. Therefore, modifications in chemical structure as per requirement of HQSAR models provide better and selective Dpp-4 inhibitors.

CONFLICT OF INTEREST

The authors confirm that this article content has no conflict of interest.

ACKNOWLEDGEMENTS

I greatly acknowledge the Council of Scientific and Industrial Research (CSIR), New Delhi for providing me the financial assistance (CSIR-SRF) during the pursuit of research goals of this work. The Authors are grateful to the Management of Institute of pharmacy, Nirma University for providing facilities and recourses for these studies.

REFERENCES

- [1] World Health Organization :Country and regional data on diabetes http://www.who.int/diabetes/facts/world_figures/en/ (Accessed May 3, 2012).
- [2] Morral, N. Novel targets and therapeutic strategies for type 2 diabetes, *TRENDS in Endo. Meta.* **2003**, *14*(4), 169-175.
- [3] Drucker, D. J. Dipeptidyl Peptidase-4 in hibition and the treatment of Type 2 Diabetes. *Diabetes Care*, **2007**, *30*(6), 1335-1343.
- [4] Holst, J.J.; Deacon, C. F. Glucagon-like peptide-1 mediates the therapeutic actions of DPP-IV inhibitors. *Diabetologia*, **2005**, *48*(4), 612-615.
- [5] Verspohl, E.J. Novel therapeutics for type 2 diabetes: Incretin hormone mimetics (glucagon-like peptide-1 receptor agonists) and dipeptidyl peptidase-4 inhibitors. *Pharma. & Therap.*, **2009**, *124*(1), 113-138.
- [6] Doupis, J.; Veve, A., DPP4 Inhibitors: a new approach in diabetes treatments. *Adv. Ther.*, **2008**, *25*(7), 627-643.
- [7] Sedo, A.; Malik, R. Dipeptidyl peptidase IV-like molecules: homologous proteins or homologous activities. *Biochimica et Biophysica Acta*, **2001**, *1550*(2), 107-116.
- [8] Lankas, G.R.; Leiting, B.; Roy, R.S.; Eiermann, G.J.; Beconi, M.G.; Biftu, T.; Chan, C.C.; Edmondson, S.; Feeney, W. P.; He, H.; Ippolito, D.E.; Kim, D.; Lyons, K. A.; Ok H. O.; Patel, R.A.; Petrov, A.N.; Pryor, K.A.; Qian, X.; Reigle, L.; Woods, A.; Wu, J.K.; Zaller, D.; Zhang, X.; Zhu, L.; Weber, A.E.; Thornberry, N.A. Dipeptidyl Peptidase IV Inhibition for the Treatment of Type 2 Diabetes. *Diabetes*, **2005**, *54*(10), 2988-2994.
- [9] Kirby, M.; Denise, M.T.; O'connor, S. P.; Gorrell, M. D. Inhibitor selectivity in the clinical application of dipeptidyl peptidase-4 inhibition. *Clinical Science*, **2010**, *118*(1), 31-41.
- [10] Myint, K. Z.; Xie, X. Q. Recent advances in fragment-based sar and multi-dimensional qsar methods. *Int. J. Mol. Sci.*, **2010**, *11*(10), 3846-3866.
- [11] Tong, W.; Lowis, D. R.; Perkins, R.; Chen, Y.; Welsh, W. j.; Goddette, D. W.; Heritage, T. W.; Sheehan, D. M. Evaluation of quantitative structure-activity relationship methods for large-scale prediction of chemicals binding to the estrogen receptor. *J. Chem. Inf. Comput. Sci.*, **1998**, *38*, 669-677.
- [12] Castilho, M. S.; Postigo, M. P.; de Paula, C. B. V.; Montanari, C.A.; Oliva, G.; Andricopulo, A. D. Two- and three-dimensional quantitative structure-activity relationships for a series of purine nucleoside phosphorylase inhibitors. *Bioorg. Med. Chem.*, **2006**, *14*(2), 516-527.
- [13] Salum, L. B.; Polikarpov, I.; Andricopulo, A. D. Structural and chemical basis for enhanced affinity and potency for a large series of estrogen receptor ligands: 2D and 3D QSAR studies. *J. Mol. Graphics Modell.*, **2007**, *26*(2), 434-442.
- [14] HQSAR™ Manual, SYBYL 7.2, Tripos Inc., St. Louis, MO. **2003**.
- [15] Anderson, A. C. The Process of Structure-Based Drug Design, *Chemistry & Biology*. **2003**, *10*(9), 787-797.
- [16] Sattigeri, J.A. Discovery of conformationally rigid 3-azabicyclo[3.1.0]hexane-derived dipeptidyl peptidase-IV inhibitors. *Bioorg. Med. Chem. Lett.*, **2008**, *18*(14), 4087-4091.
- [17] Doddareddy, M.R.; Cho, Y.S.; Koh, H.Y.; Pae, A.N. CoMFA and CoMSIA 3D QSAR analysis on N1-arylsulfonylindole compounds as 5-HT6 antagonists. *Bioorg. Med. Chem.*, **2004**, *12*, 3977-3985.
- [18] Cavalcanti, A.R.O.; Leite, E.S.; Neto, B.B.; Ferreira, R. On the classes of aminoacyl-tRNA synthetases, amino acids and the genetic code. *Origins Life Evol. B*, **2004**, *34*(4), 407-420.
- [19] Wang, K.L.; Wen, Z.N.; Nie, F.S.; Li, M.L. A new hybrid model of amino acid substitution for protein functional classification. *Chin. Chem. Lett.*, **2005**, *16*, 1133-1136.
- [20] Glide, version 4.5, Schrödinger, LLC, New York, NY, **2007**.
- [21] LigPrep-v2.1, Molecular Modeling Interface, Schrödinger LLC, NY, **2007**.
- [22] Metropolis, N.; Rosenbluth, A.W.; Rosenbluth, M.N.; Teller, A.H.; Teller. Equation of State Calculations by Fast Computing Machines. *E. J. Chem. Phys.*, **1953**, *21*(6), 1087-1092.