Original Article

Journal of Pharmaceutical and Medicinal Chemistry
Volume 2 Number 2, July - December 2016
DOI: http://dx.doi.org/10.21088/jpmc.2395.6615.2216.1

2D QSAR Study on Some Non-Steroidal 4-(Anilino) Pyrrole-2-Carboxamides as Androgen Antagonists

Sarabjit Kaur*, Richa*, Renu Chadha*, Neelima Dhingra*

(Received on 21.06.2016**, Accepted on** 25.06.2016**)**

Abstract

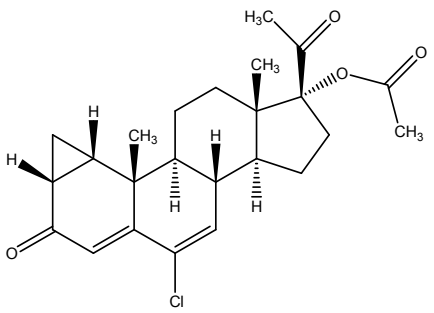
In this paper, an attempt has been made to develop a quantitative structure–activity relationship 2D-QSAR study on 30 analogues of 4-(anilino) pyrrole-2-carboxamide, non steroidal non anilide molecules as androgen antagonist, to simulate a mathematical relation and to estimate the contributing descriptors. A statistically validated 2D QSAR model was obtained through k-nearest neighbour (kNN) simulated annealing with statistical data (q^2=0.792; pred r^2 = 0.862) method using Vlife molecular design suits (MDS) 4.3. Positive and negative correlation of the various descriptors have been discussed.

Keywords: Androgen Antagonists; Bicalutamide; Flutamide; Non Steroidal; Prostate Cancer; 2D-QSAR.

Introduction

Prostate cancer (PCa) is the most enigmatic of the common solid malignancies and second second-leading cause of cancer death in men, beyond middle age [1,2]. The human prostate is a hormone sensitive organ and normal growth development and maintenance of the prostate depends on androgens such as testosterone (T) and dihydrotestosterone (DHT), acting through the androgen receptors (AR). The AR, also known as NR3C4 (nuclear receptor subfamily 3, group C, member 4), is a type of nuclear receptor and activated by binding of these androgens in the cytoplasm, followed by their translocation into the nucleus [3]. The regulation of androgen biosynthesis or its action on the androgen receptor is central to the management of prostate cancer [4]. AR has been found to be important in the development and progression of prostate cancer, and androgen deprivation by preventing binding of androgens to the AR represents one of the mainstay interventions for the treatment of androgen-dependent PCa. Typical AR antagonists antagonize the biological responses elicited by endogenous and/or exogenous androgens by competitively inhibiting the binding of the latter to AR [5]. During the last two decades a number of non-steroidal [6] and steroidal compounds [7,8] have been prepared as competitive or non-competitive androgen antagonist and of which some, such as cyproterone acetate (1), flutamide (2) and bicalutamide (3) are clinically used to treat PCa [9-12].

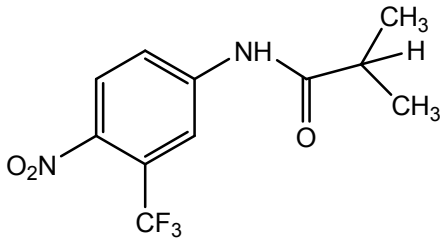
(1)



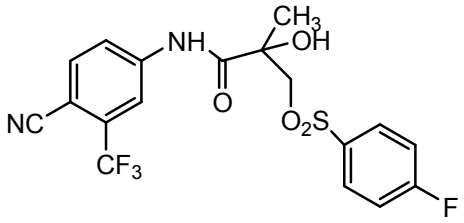
Author Affiliation: *University Institute of Pharmaceutical Sciences, Punjab University, Chandigarh.

Reprint Request: Neelima Dhingra, Assistant Professor, University Institute of Pharmaceutical Sciences, Department of Pharmaceutical Chemistry, Panjab University Chandigarh-160014.

E-mail: neelimad08@gmail.com



(2)



(3)

Recent second generation AR antagonists have been designed that retain antagonism in over-expressing cell lines, and among these agents enzalutamide has recently successfully met efficacy criteria in a large Phase III clinical trial [13]. With this knowledge, it is reasonable to suggest that effective strategies (investigational new drugs) that lead to AR down-regulation may be useful for preventing the development, progression and treatment of PCa.

Computer aided drug designing based on Quantitative Structure Activity Relationship (QSAR) between biological activity and physicochemical descriptor, is a tool that has been used to increase the efficiency of the drug discovery process. Though, QSAR does not completely eliminate the trial and error factor involved in the development of a new drug, but certainly it decreases the number of compounds synthesized by facilitating the selection of the most promising examples [14]. Different statistical approaches are used while building a QSAR model, and these models can be used to predict the activity of new compounds. Considering the recent interest in androgen receptors in the progression of PCa and further designing and development of AR antagonists, a QSAR investigation of the aforementioned series is carried out using two dimensional (2D) molecular descriptors. The present study aimed to establish relationship AR antagonist for structurally related 4-(anilino) pyrrole-2-carboxamide, derivatives and the physicochemical descriptors in quantitative terms. The statistically validated two dimensional (2D QSAR) model was obtained through k-nearest neighbour (kNN) analysis method using Vlife molecular design suits (MDS).

Material and Method

Biological Activity and Molecular Descriptors

Cytotoxicity-inducing activity data IC_{50} (μM) of the molecules (Table 1, 2) were taken from the published reports [5]. The experimental IC_{50} values were evaluated using Shionogi carcinoma-3 (SC-3) cells, IC_{50} representing 50% inhibition of this enzyme. The negative logarithm of the measured IC_{50} (μM) [pIC_{50} = -log (IC_{50})] was used as dependent variable for 2D studies and it is listed in Table 3. Two-dimensional (2D) structures were constructed using Chemdraw Ultra 8.0, and then converted to three-dimensional (3D) structures in same software. All 3D molecules were subjected to energy minimization using molecular mechanics (MMFF) and conformational analysis using Montocarlo conformational search until the root mean square (RMS) gradient value reaches a value smaller than 0.001kcal/mol Å. This Montocarlo conformational search method generates various conformations of the stable molecules after energy minimization and calculates its physicochemical descriptors. Energy-minimized geometry was used for calculation of descriptors, a total of 408 2D descriptors were calculated, eventually reduced to 222 after applying invariable column selection. The steric, topological, electronic, thermodynamic, molecular, and structural descriptors calculated were consists of chiV6chain, chiV4pathcluster, IdwAverage (steric), chi1, chi5, chiV0, chiV3, chiV4, chiV5, 3PathCount, chi6chain, chiV6chain, chiV3Cluster, 3ClusterCount, chi4pathCluster, chiV4pathCluster, 4path Cluster Count, kappa3, k1alpha, k2alpha, VChi 3 cluster , VChi 4 cluster, VChi 5 path, Kier shape 2, Kier alpha 1, Kier alpha 3, Kier symmetry index, Chi 0, Chi 2, Chi 3 cluster, Chi 5 path, VChi 1, VChi 3 path, VChi 4 path, VChi 4 path/ cluster, Kier shape 1, Kier shape 3, Kier alpha 2, Kier flexibility, Kier (steric), charge index 8, valence charge index 10, bound charge index 2, k2alpha, Chi 5 path, 0PathCount, (topological), Max. positive charge, max. positive hydrogen charge, local dipole index, relative negative charge, -ve potential, average potential, average +ve potential, average -ve potential, most +ve & -ve potential distance, average potential, average +ve potential, average -ve potential (electronic). Besides these all, alignment independent descriptors were also calculated.

Selection of Training and Test Set

A data set of 30 molecules belonging to 4-(anilino) pyrrole-2-carboxamides as novel androgen receptor (AR) antagonists were taken from the literature [5] and used for QSAR study. The dataset was divided into training set (21 compounds) and test set (9 compounds) by Random selection(RS) method for principal component regression (PCR), multiple linear regression (MLR), partial least squares (PLS) and k-Nearest Neighbour (kNN). In random selection the test set molecules are selected randomly from the whole data set. Inhibitory activity i.e. pIC_{50} has been considered as dependent variable and the remaining descriptors as independent parameters.

Regression Analysis

QSAR is a parametric approach applied to set of series of compounds in order to understand their mechanism of action and important contributory structural factors responsible for activity. In this regard selected physicochemical or structural parameters were used as the correlative parameters and related to the observed biological activity of the 30 molecules by regression analysis PCR, MLR, PLS and kNN as statistical approaches. The cross-correlation limit was set at 0.7, and term selection criteria as r^2, *F*-test 'in,' at 4 and 'out' at 3.99, r^2, and *F*-test. Variance cut off was set at 0.1, scaling to auto scaling, and number of random iterations to 10. Statistical measures were used for the evaluation of QSAR models were the number of compounds in regression *n*, multiple correlation coefficient (r), coefficient of determination (r^2), number of descriptors in a model *k*, *F*-test (Fisher test value) for statistical significance *F*, cross-validated correlation coefficient q^2, predictive squared correlation coefficients pred_r^2, coefficient of correlation of predicted data set pred_r^2se and standard error (SE) of estimation r^2se and q^2se. The regression coefficient r^2 is a relative measure of fit by the regression equation. The correlation coefficient values must be closer to 1.0 that represents the better fit of the regression. The ratio of the variance explained by the model and the variance due to the error in the regression is being reflected by F test and high value further indicates the statistically significant model. Predictive r^2 (r^2_pred) is the one of the validation parameter, that was calculated for evaluating the predictive capacity of the model and if its value greater than 0.5, means of the QSAR model has good predictive capacity. The number of statistical models were developed using PCR, MLR, PLS and kNN based regression methods coupled with forward, genetic algorithm simulated annealing and forward backward method and correlated the biological activity with the physico-chemical descriptor values.

Validation of QSAR Model

Model validation is done to analyse the internal stability and predictive ability of the QSAR models. The best to evaluate quality of regression model is internal and external validation. Internal validation is carried out using leave one out (q^2, LOO) method. For calculating q^2 each molecule in the training set was eliminated once from training set and the activity of eliminated molecule was predicted by using model developed by the remaining molecules. This q^2 described the internal stability of the model [55] and is expressed as shown in equation (Eq A.1)

$$q^2 = 1 - \Sigma (Y \text{ pred} - Y \text{ act})^2 / \Sigma (Y \text{ act} - Y \text{ mean}) \quad \text{(Eq A.1)}$$

Secondly, the predictive ability of the model i.e. external validation was confirmed by comparing the observed value of the test set molecules with predictive value of test molecules and is indicated by predicted r^2 as shown in equation (Eq A.2)

$$\text{pred\_}r^2 = 1 - \Sigma (Y \text{ pred(Test)} - Y \text{ Test})^2 / \Sigma (Y \text{ Test} - Y \text{ Training}) \quad \text{(Eq A.2)}$$

Results and Discussion

With the aim to establish predictive model, which could be further utilized for the designing and synthesis of androgen receptor (AR) antagonists, PCR, MLR, PLS and kNN methods were employed to relate the corresponding inhibitory activities i.e. IC_{50} values with the selected computed molecular parameters for all the 30 molecules. Model having best fit with minimum number of descriptors was found to be the best model. When this point is achieved, no further considerable improvement in the regression coefficient (r^2 and q^2) values were observed, even if a new descriptor is added. In the present study, 250 equations were generated employing three; four or more variable combinations for combined dataset, and the model with good statistical values for better correlation have been given in the Table 4. The kNN methodology relies on a simple distance learning approach, whereby an unknown member is classified according to the majority of its k-nearest neighbours in the training set. The nearness is measured by an appropriate distance metric (e.g., a molecular similarity measure calculated using field interactions of molecular structures) [15].

➢ *K-Nearest Neighbour- Genetic Algorithms (kNN-GA)*

This method employs a stochastic variable selection procedure, combined with kNN, to optimize (i) the number of nearest neighbours (k) and (ii) the selection of variables from the original pool as described in simulated annealing [15].

➢ *K-Nearest Neighbour- Stepwise-Forward (KNN-SW-F)*

This method employs a stepwise variable selection procedure combined with kNN to optimize (i) the number of nearest neighbours (k) and (ii) the selection of variables from the original pool as described in simulated annealing. The step by-step search procedure begins by developing a trial model with a single independent variable and adds independent variables, one step at a time, examining the fit of the model at each step. The method continues until there are no more significant variables remaining outside the model [15].

➢ *K-Nearest -Neighbour- Simulated Annealing (KNN-SA)*

kNN-SA is the simulation of a physical process, 'annealing', which involves heating the system to a high temperature and then gradually cooling it down to a preset temperature (e.g., room temperature). During this process, the system samples possible configurations distributed according to the Boltzmann distribution so that at equilibrium, low energy states are the most populated [15].

kNN-SA (model 1) was found to be the best model and exhibits good external predictivity indicated by pred r^2 0.86, whereas squared cross correlation coefficient was found to be q^2 0.79. The fitness plot, contribution plot and actual and predicted activities of training and test set molecules of the 30 molecules have been given in (Figure 1) and (Figure 2a & 2b) respectively. Table 5 describes the various descriptors, their categories as well as their percentage contribution towards good model. The best model is indicating the effect of descriptors like SsOHE index, SaaCHcount, SaasCcount, kappa1, T_C_C_5 on biological activity. It is also evident from the statistical data that T_C_C_5count is a topological descriptor lies within the range between (8.00-23.00), that is correlated with biological activity, that means upto increase in 23% of T_C_C_5count will increase the androgen antagonistic activity upon human prostate tumor. On other hand, SaaCHcount should lie in range of 3.00- 6.00, that means the increase upto 6% in the number of carbon atoms connected with a hydrogen along with two aromatic bonds, which will increase the inhibitory activity against human prostate tumor. SsOHE index is a topological descriptor which indicates the number of –OH group connected with one single bond that must be increased upto 9.3%. SaasCcount is a topological descriptor that lie in range between 3.00-5.00, which indicates the number of carbon connected with one single bond along with two aromatic bonds, that will increase androgen antagonistic activity. kappa1 is a topological descriptor which signifies first kappa shape index it must lies in the range of 15.87-19.75, for good androgen antagonistic activity upon human prostate tumor.

In conclusion we have developed quantitative structure–activity relationship (QSAR) models of 4-

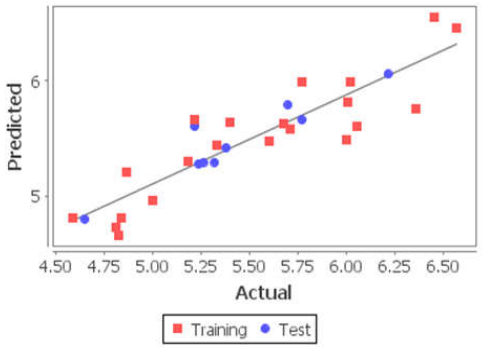
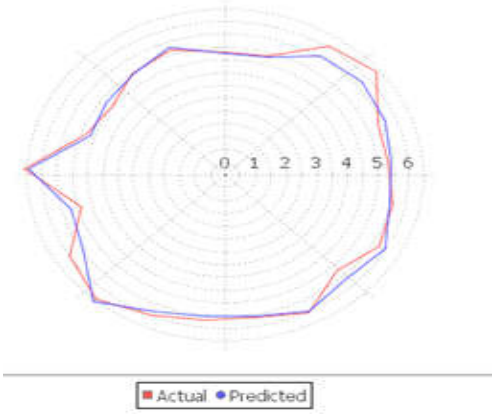


Fig. 1: Fitness plot of kNN-SA



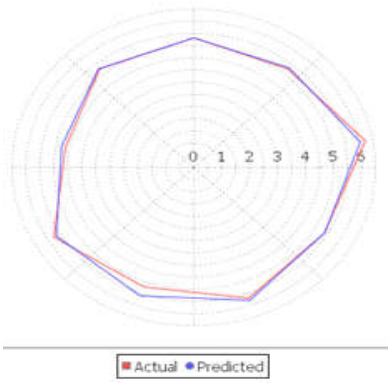
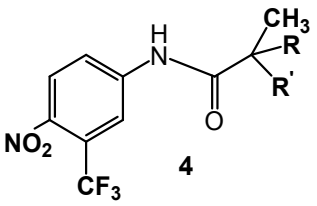


Fig. 2a, 2b: Actual and Predicted activity of training set and test molecules by kNN-SA

Table 1: Series of 4-(Anilino) pyrrole-2-carboxamides



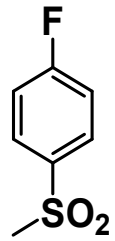
4

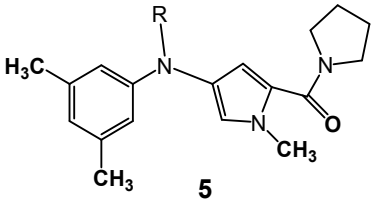
R= **R′=**

(a) H CH_3

(b) OH CH_3

(c) OH





5

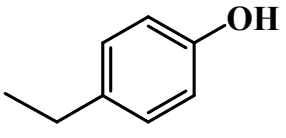
R=

(a) H

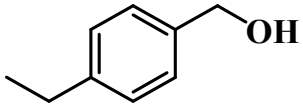
(b) CH_3

(c) C_2H_5

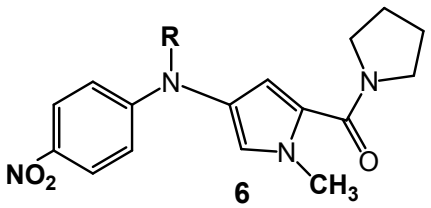
(d) $CH(CH_3)_2$



(e)



(f)

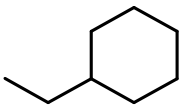


6

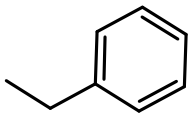
R=

(a) CH_2CH_3

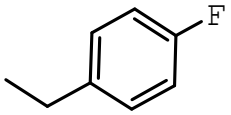
(b) $CH(CH_3)_2$



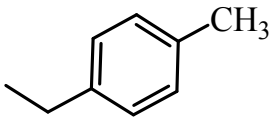
(c)



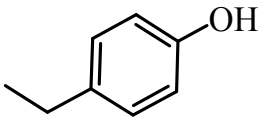
(d)



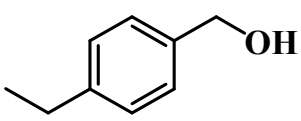
(e)



(f)

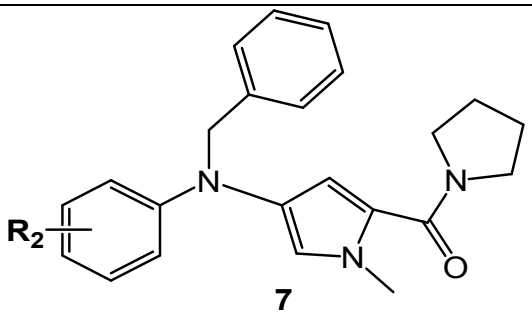


(g)



(h)

Table 2: Series of 4-(Anilino) pyrrole-2-carboxamides



7

R=

(a) 4-NO_2

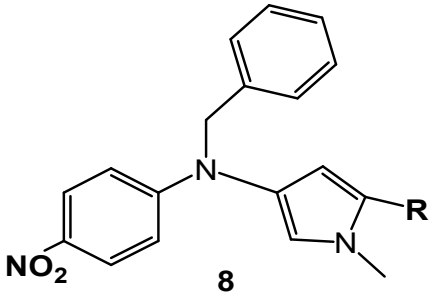
(b) 3-CF_3, 4-NO_2

(c) 4-CN

(d) 4-CF_3

(e) 4-$COCH_3$

(f) 4-$COOCH_3$



8

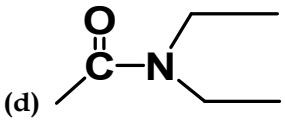
R=

(a) $COOC_2H_5$

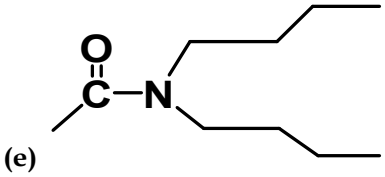
(b) CH_2OH

(c) $CONH_2$

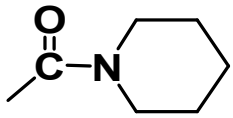
(d)



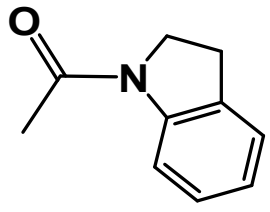
(e)



(f)



(g)



(h)

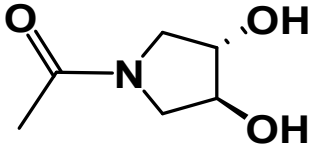


Table 3: Observed and Predicted activities of compounds

Compound No.	Observed Activity (IC_{50})	Predicted Activity(IC_{50})
4a[t]	5.3767	4.73033
4b	6.4559	4.81222
4c	6.5686	5.60073
5a	4.8239	5.75299
5b	4.585	5.63742
5c[t]	4.6478	5.44405
5d	5	5.29439
5e	5.1804	5.29569
5f[t]	5.2365	5.47549
6a	4.8096	6.05795
6b	4.8632	5.42217
6c	6.0555	5.99330
6d	6.3565	5.79186
6e	5.3979	5.65963
6f	5.3279	5.99285
6g[t]	5.3187	5.62476
6h[t]	5.2596	5.57752
7a	5.602	5.81876
7b	6.2146	5.60750
7c[t]	5.7695	5.66017
7d	5.6989	5.48387
7e[t]	5.2146	6.55692
7f	6.0222	5.21131
8a	5.6777	6.45301
8b	5.7099	4.66519
8c	6.0043	4.81721
8d	5.2146	4.80262
8e[t]	5.7695	4.96899
8f[t]	6	5.30619
8g	4.8632	5.28166

Table 4: Model obtained by kNN methods and its statistical parameters

Models	q^2	Pred r^2
kNN -SA	0.79	0.86

Table 5: Various descriptors, their categories as well as their contribution towards

Descriptor	Category	Contribution range	Meaning
kappa1	Topological	15.87-19.75	This descriptor signifies first kappa shape index: (n-1)2 / m2
SsOHE-index	Topological	0.00-9.37	Number of –OH group connected with one single bond.
T_C_C_5	Topological	8.00-23.00	Number of Carbon atoms (single double or triple bonded) separated from another carbon atom (single or double bonded) by 1 bond distance in a molecule.
SaaCHcount	Topological	3.00-6.00	This descriptor defines the total number of carbon atoms connected with a hydrogen along with two aromatic bonds
SaasCcount	Topological	3.00-5.00	This descriptor defines the total number of carbon connected with one single bond along with two aromatic bonds.

(anilino) pyrrole-2-carboxamide derivatives as androgen antagonists for the potential treatment of prostate cancer. After splitting the datasets into training and test sets, the respective descriptors were selected from the pool of multiple descriptors using the kNN linear regression method and the most relevant descriptors were selected to build the model. We expect that these investigations will further help

in rationalizing the design and synthesis of more potential androgen antagonist anticancer agents.

References

1. Simon C, David R, Declan C, Alejo RV, Lars H, Henrik M. Causes of death in men with prostate cancer: an analysis of 50 000 men from the Thames Cancer Registry BJU International. 112(2): 182-89.
2. Craig J, Newschaffer, Koichiro O, M. Kathleen McD, Lynne TP. Causes of Death in Elderly Prostate Cancer Patients and in a Comparison Nonprostate Cancer Cohort. Journal of National Cancer Institute 2000; 92(8): 613-21.
3. Freedman LP. Molecular Biology of Steroid and Nuclear Hormone Receptors; Birkhauser: Boston, MA, 1998.
4. Van BA, Varella-Garcia M, Korch C, Johannes WU, Smith EE, *et al*. Molecular characterization of human prostate carcinoma cell lines. Prostate. 2003; 57: 205-22.
5. Ken-ichi W, Keisuke I, Hiroyuki M, Yuichi H, Aya T. 4-(Anilino) pyrrole-2-carboxamides: Novel non-steroidal/non-anilide type androgen antagonists effective upon human prostate tumor LNCaP cells with mutated nuclear androgen receptor. Bioorganic and Medicinal Chemistry. 2008; 16: 6799–6812.
6. Negro VA. Selective androgen receptor modulators (SARMs): a novel approach to androgen therapy for the new millennium. Journal of Clinical and Endocrinology Metabolism. 1999; 84: 3459–62.
7. Gao, W, Bohl C E, Dalton JT. Chemistry and structural biology of androgen receptor. Chemical Reviews. 2005; 105: 3352–70.
8. Gao WQ, Dalton JT. Ockham's razor and selective androgen receptor modulators. via selective androgen receptor modulators (SARMs) Drug Discovery Today. 2007; 12: 241–8.
9. Neri RO. Antiandrogens. Advances in Sex Hormone Research Journal. 1976; 2: 233–62.
10. Maucher A, Von AE. Antiproliferative activity of casodex (ICI 176: 334) in hormone-dependent tumours. Journal of Cancer Research and Clinical Oncology. 1993; 119: 669–74.
11. Sufrin G, Coffey DS. Flutamide, mechanism of action of a new nonsteroidal antiandrogens. Investigative Urology Journal. 1976; 13: 429–34.
12. Narayana, AS, Loening SA, Culp DA. Flutamide in the treatment of metastatic carcinoma of the prostate. British Journal of Urology. 1981; 53: 152–53.
13. Beer TM, Armstrong AJ, Rathkopf DE, Loriot Y, Sternberg CN, Higano CS *et al*., Enzalutamide in Metastatic Prostate Cancer before Chemotherapy, The New England Journal of Medicine. 2014; 371-5.
14. Kapetanovic IM. Computer-aided Drug Discovery and Development (Caddd): *in silico*-chemico-biological approach. Chemico- Biological Interactions. 2008; 171: 165-8.
15. Malleshappa NN, Harun MP. A comparative QSAR analysis and molecular docking studies of quinazoline derivatives as tyrosine kinase (EGFR) inhibitors: A rational approach to anticancer drug design. Journal of Saudi Chemical Society. 2013: 17: 361–79.