Research Article
Open Access

In silico Identification of potential 5α–reductase inhibitors for prostatic diseases: QSAR modelling, molecular docking, and pre ADME predictions

Abstract

Background: Benign Prostatic Hyperplasia and prostate cancer are some of the most common old age men diseases and caused by augmented level of potent androgen dihydrotestosterone (DHT). Preventing DHT synthesis via 5α–reductase (5–AR) inhibition has been shown to have a remarkable effect on prostatic disease with low toxicity. Thus, there is much interest in the potential role for 5–AR inhibitors (5–ARIs) in the management of prostatic diseases. Some of the natural products, semi synthetic derivatives especially steroidal molecules, are effective inhibitors of 5AR. However, given the limited number of clinically approved inhibitors and the associated side effects, the discovery of new inhibitors is urgent. Here, the *In silico* approaches have been performed to identify new potential inhibitors of 5–AR.

Methods: In the present study, two and three–dimensional (2D and 3D) QSAR models have been developed using a selected series of steroidal derivatives as 5ARIs, in order to elucidate the structural properties required for 5α–Reductase (5AR) inhibitory activity and anticancer activity. Further *In–silico* studies (molecular docking and pre–ADME) of selected series have also been carried out to identify the binding orientation and the receptor ligand interactions responsible for exhibited activity and the drug like properties.

Results: Best 2D–QSAR model was generated using Simulated Annealing–Partial Least Square (SA–PLS) method (r^2=0.9669, q^2=0.8720, F–value=109.4339 and pred–r^2=0.8289), whereas simulated annealing–k nearest neighbor (SA–kNN) approach provided us with best 3D–QSAR model (q^2=0.9523 and pred–r^2=0.8068). The compounds were further sorted by applying ADME properties against prostate cancer cell lines PC–3. Docking analysis indicated that the selected series of compounds have comparable binding affinity with the receptor protein and suggested that hydrophobic and electrostatic moieties can have a key role in the inhibition mechanism. These findings can provide a new strategy to develop new steroidal 5ARIs useful in prostatic disease.

Keywords: 5AR inhibitors, dihydrotestosterone, docking, QSAR, SA–kNN

Volume 2 Issue 3 - 2018

Akanksha, Vikas Mehta, Richa Dhingra, Monika, Neelima Dhingra

Department of Pharmaceutical Chemistry, Panjab University, India

Correspondence: Neelima Dhingra, Department of Pharmaceutical Chemistry, Panjab University, India, Tel +91–9876573312, Email neelimad08@gmail.com

Received: May 26, 2018 | **Published:** June 12, 2018

Introduction

Steroids assume an essential part in a several processes, ranging from the differentiation, growth, development, physiological and regenerative functions in the human body,[1] because of their ability to cross membrane easily.[2] Steroids as well as their derivatives have the potential to be developed as drugs for the treatment of a large number of diseases including cardiovascular, autoimmune diseases, brain tumours, breast cancer, osteoarthritis, prostate cancer etc.[3–6] The promise of using steroids for development of lead molecules lies in the regulation of a variety of biological processes by these molecules and being a fundamental class of signalling molecules.[7,8] Benign prostatic hyperplasia (BPH) and Prostate Cancer (PC), are the leading disorders of the old age men. These prostatic diseases are characterized by a progressive enlargement of prostatic tissue, that results in the obstruction of proximal urethra and cause urinary flow disturbances.[9] Nearly half of men aged over 50years show histological evidence of BPH and PC and proportion increases to 80% by the age of 70.[10] Dihyrotestesterone (DHT) hypothesis postulates that androgens play an important role in growth of prostate. Male hormone testosterone (T), is biosynthesized in testicles and adrenal glands, and converted to more potent androgen i.e. dihydrotestosterone (DHT) by NADPH dependent enzyme 5α–Reductase (5AR). 5AR is a membrane–bound enzyme that irreversibly catalyses the reduction of 4–ene–3–oxosteroids to the corresponding 5α–3–oxosteroids Figure 1. Three isozymes have been identified for 5AR based on their gene sequence, location, and pH. Type 1 isozyme (5AR1) is expressed in skin and liver and show maximal activity at pH ranging from 6.0~8.5, whereas type 2 isozyme (5AR2) shows its activity at pH 5.5 in prostate and other genital tissues.[11] Recently, type 3 isozyme (5AR3) has also been identified in castration–resistant prostate cancer cells including other tissues like pancreas, brain, skin and adipose tissues.[12–14] 5AR has emerged as a therapeutic target, as 5α–Reductase Inhibitors (5ARIs) by causing the suppression of DHT biosynthesis provided a logical treatment for BPH and PC.[15,16] Further, isolation and characterisation of the three different isozymes have led in the advancement of new and selective inhibitors with improved anticancer therapies.[17] Computer Aided Drug Designing (CADD) in perspective of QSAR between natural activity and physicochemical descriptor, is a gadget that has been utilized to build the productivity of the medication disclosure process.[18] Though primary sequence of 5AR isozymes, are available, but crystal structure of 5AR have yet not been determined as yet they have not been isolated and purified from tissues or cells despite. In the absence of structural information of target protein, different receptor mapping techniques can create permitting developing 3D surrogate of the coupling pocket and could be utilized to predict

©2018 Akanksha et al. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and build upon your work non-commercially.

the binding interactions and energy. In present study, availability of the several crystal structures of 5β–reductase (5BR) with various steroids and same cofactors (NADP or NADPH) have been utilized to identify 5ARIs by performing virtual screening the series of compound. Statistically validated QSAR model were developed using 16–dehydropregnenolone acetate derivatives(1a–1e),[19,20] 21–imidazolyl–16–dehydropregnenolone derivatives (2a–2e),[21] androstano [17,16–d][1,2,4] triazolo [1,5–a] pyrimidine derivative and Testosterone–imidazo [1,2–α] pyridine hybrids (4a–4b), 4–(aryl)–4–pregneno [3,2–e] pyridinone derivatives (5a–5h, 6a–6h) in order to identify important physicochemical parameters, followed by molecular docking. Binding free energy was estimated to interpret the stability and interaction of 5AR compound complexes. In addition, pre ADME i.e. absorption, distribution, metabolism, and excretion have been predicted to estimate basic pharmacological properties. The observation of the present study can provide an important measurement and perspective to better understand the inhibitory mechanism of 5AR for future prostate therapies.[22,23]

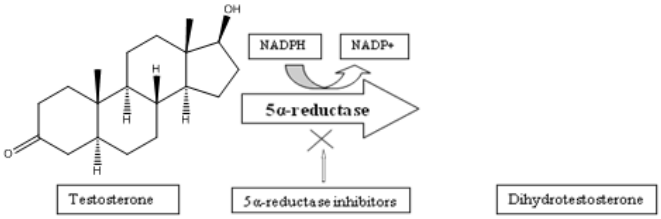


Figure 1 Conversion of Testosterone (T) to Dihydrotestosterone (DHT) by 5α–Reductase and inhibition of 5α–Reductase by 5α–reductase inhibitors.

Experimental methods

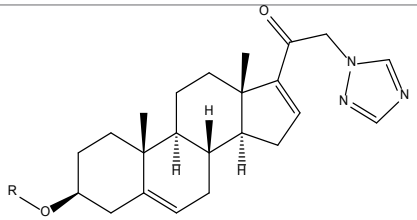
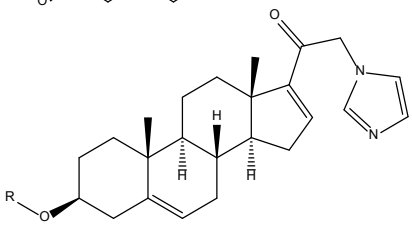
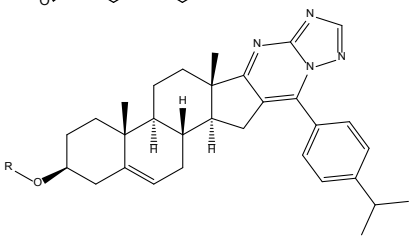
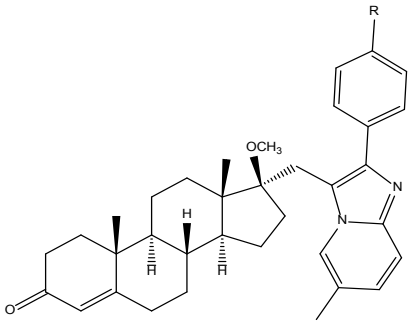
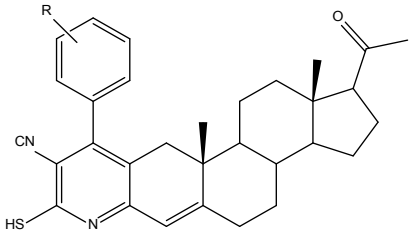
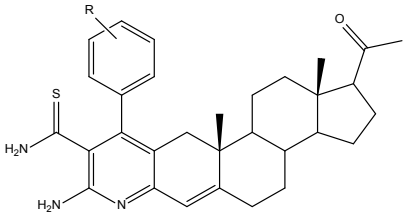
29 molecules belonging to various pregnenolone and testosterone series as 5ARIs showing variance in their structure and potency were taken Table 1 and used for QSAR and *in–silico* studies. All molecular modelling studies (2D and 3D) were performed using the Vlife MDS 4.6 QSAR plus software on an HP Pentium IV 2.80 GHz Processor/ Microsoft Win XP Home Edition system. Docking studies were carried out using BIOPREDICTA tool available in Vlife MDS software. The structures of all compounds were sketched in Chem Draw Ultra 12.0. All structures are cleaned and optimized. Energy minimization of the molecules was conducted using the Merck molecular force field method with the root mean square gradient set to 0.01 kcal/molÅ, and the iteration limit to 10,000. The conformers for all structures are generated and selected the low energy conformer for each compound and used for further study.

2D–QSAR study

Dataset and molecular modelling for 2D–QSAR

Cytotoxicity–inducing activity data IC_{50} (μM) of the molecules Table 2 were taken from the published reports. The experimental IC_{50} values were evaluated by Serum Bactericidal Rate (SBR) assay method, using PC cell lines (PC–3).[24] The negative logarithm of the measured IC_{50} (μM) [pIC_{50}=–log (IC_{50})] was used as dependent variable for 2D study. The total set of 29 compounds was divided into a training set (20 compounds) and test set (5 compounds) for generating 2D–QSAR models and an validation test set of 10% compounds (4 compounds) for validating the quality of the models. Important and the key feature of any QSAR model is the selection of molecules for training and test set and care was taken retain the molecules in test with maximum and minimum value range of biological activities of training set of compounds.

Table 1 Series of various selected derivatives as 5–ARIs

Compound	Structure	R
1a		$COCH_3$
1b		COC_3H_7
1c		COC_4H_9
1d		COC_5H_{11}
1e		H
2a		H
2b		$COCH_3$
2c		COC_2H_5
2d		COC_4H_9
2e		COC_5H_{11}
3		H
4a		CH_3
4b		C_2H_5
5a		H
5b		4–Br
5c		2–Cl
5d		4–Cl
5e		4–F
5f		2–NO_2
5g		4–NO_2
5h		4–OCH_3
		H
6b		4–Br
6c		2–Cl
6d		4–Cl
6a		
6e		4–F
6f		2–NO_2
6g		4–NO_2
6h		4–OCH_3

Copyright: ©2018 Akanksha et al.

Table 2 Observed and predicted activity of the selected series of compounds

Compound	Observed activity	Predicted activity
1a	4.25	4.242
1b	5.036	4.94
1c	4.417	4.386
1d	4.772	5.086
1e	4.212	4.262
2a	4.016	3.947
2b	4.099	4.167
2c	4.645	4.589
2d	4.625	4.706
2e	4.764	4.897
3	5.137	5.133
4a	4.958	4.926
4b	4.931	5.025
5a	5.02	4.964
5b	4.976	5.051
5c	5.004	5.078
5d	4.98	5.049
5e	5.009	5.028
5f	5.115	4.949
5g	5.127	5.031
5h	4.971	5.014
6a	4.937	4.883
6b	4.857	4.936
6c	4.901	5.04
6d	4.893	4.913
6e	4.924	4.847
6f	4.949	4.965
6g	4.967	4.933
6h	4.847	4.847

Descriptor calculation

Its well know that the drug or the compounds always bind to effectors/receptors in the most stable form i.e. the minimum energy form. Present study involves the involves the calculation of theoretical individual molecular descriptors (steric, topological, electronic, thermodynamic, molecular, and structural) for geometrically optimized structures such as molecular weight, volume, XlogP; molecular refractivity (smr); estate numbers/contributions, polar surface area, element count, dipole moment, hydrophobicity XlogpA/SlogpA; topological such as T_2_Cl_6, T_C_Cl_6, T_T_S_7, T_T_Cl_7 with a view to develop structure–activity relationship of the selected pregnenolone and testosterone derivatives against the 5AR enzyme. Total of 240 2D descriptors were calculated and were eventually reduced to 175 after applying invariable column selection. Different regression methods such as Principal Component Regression (PCR), Multiple Linear Regression (MLR) and Partial Least Square (PLS) were employed for building the models and their validation and robustness was verified by an external test set of 10% compounds.[25]

Statistical analysis

The number of statistical models were developed using MLR, PCR and PLS based regression methods coupled with forward, forward backward, genetic algorithm (GA) and simulated annealing (SA) method and correlated the biological activity with the physico–chemical descriptor values. The program computes the best model on the basis of squared correlation coefficient r^2, crossed validated q^2 (relative measure of quality of fit); Fischer's value F–test (between the variance of calculated and observed activity and pred_r^2),. Comparison of calculated value of F–test with tabulated value of F–test shows the level of statistical significance (99.99%) of the QSAR model. Absolute quality of fitness of the model is being determined by low standard error of pred_r^2 se, q^2_se, and r^2_se. The generated QSAR model is generally validated for predictive ability inside the model by using cross validation (leave–one–out–LOO) for q^2; followed by external validation i.e. more robust alternative method, which divides the data into training set & test set to calculate pred_r^2. The high pred_r^2 and low pred_r^2 se indicates the high predictive ability of the model. The statistical significance of selected 2D–QSAR model are further supported by the 'fitness plot' (observed versus predicted activity) and provides an idea about how well the model is trained and how well it can predicts the activity of the external test set. The contribution chart for the significant model gives the percentage contribution of the descriptors used in deriving the model.

3D–QSAR study

Data set and molecular modelling for 3D–QSAR

The total set of 29 compounds was divided into a training set (20 compounds), test set (5 compounds) for generating 3D–QSAR models and a validation test set consisting of 4 compounds i.e 10% compounds of the total. Optimal sets were generated using the sphere exclusion (SE) algorithm with dissimilarity value of 13.4. The SE method employs the following algorithm: (i) selection a point and its inclusion in the training set; (ii) building a sphere with radius R and a center; (iii) inclusion of all the points within the sphere, except for the center, in the test set; (iv) discarding all points in the sphere from the initial set; (v) if no points are left, stop, otherwise repeat whole process starting from step (i). The most active compound in the dataset is selected as the starting point for building a sphere.

Alignment procedure

The very crucial and sensitive step of the 3D–QSAR study is molecular alignment of energy minimized and geometry optimized structure in 3D space against the template. In general, geometric similarity should exist between the modelled structures and the bioactive conformation of the different molecules for good and reliable 3D–QSAR model. The alignments not only relates the conformation flexibility but also define the putative pharmacophore for the series of ligands. Alignment of all 29 compounds was done using the Vlife MDS 4.6 template based alignment tool, where the most active molecule was taken as a reference and 17–substituted 5–androsten–3β–ol served as a template in MDS Figure 2. Multiple conformation of each molecule was generated using Monte Carlo conformation search method.[26]

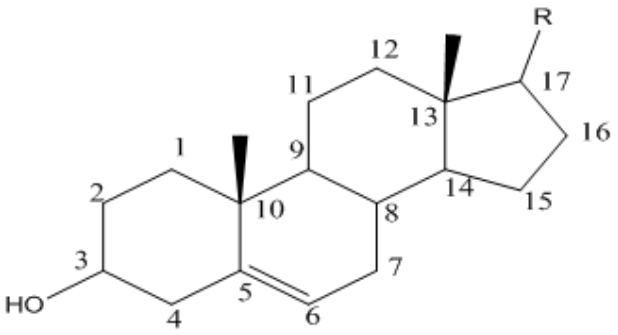


Figure 2 Template for the alignment of the data set.

Descriptors calculation

Molecular or physicochemical descriptors were calculated of all the biologically active aligned conformations of selected series of molecules after their energy minimization using Vlife MDS 4.6 which allows the user to choose probe, grid size and grid interval for the generation of descriptors. The dielectric constant was set to 1.0, considering the distance–dependent dielectric function probe setting was carbon atom with charge 1.0. This resulted in calculation of various descriptors like electrostatic, steric, and hydrophobic field for all the compounds in separate columns. Invariable columns do not contribute to the development of QSAR model thus are removed before proceeding. 3D–QSAR studies were carried out by SA–kNN method as variable selection method as follows:

a. Distances were calculated between an unknown object & all the objects in the training set.

b. Numbers of nearest neighbours (k) were selected from the training set, according to the calculated distances.

c. Classification of the object with the group to which the majority of the k objects belong, followed by selection of an optimal k value.

The variables and optimal k values were chosen using stepwise variable selection method in order to optimize.

The number of nearest neighbours (k) and Selection of variables from the original pool. Initially a trial model is developed with a single independent variable and adds independent variables, one step at a time followed by examination of the fit of the model at each step (using weighted kNN cross–validation procedure). The process continues till no more significant variable remained outside the model and once the training and test sets are generated, kNN methodology is applied to the descriptors generated over the grid. The steric, electrostatic, and hydrophobic energies are computed at the lattice points of the grid using a methyl probe of charge +1.

Model validation

Internal validation

The internal validation was done by implementing standard LOO procedure, wherein a molecule a molecule is eliminate from the training set its biological activity is predicted as the weighted average activity of the most similar molecules. This step is repeated until all the molecules from the training set have been eliminated for their activity prediction. The cross–validation r^2 (q^2) value was also calculated using Eq A.1 where y_i and $\hat{y}_i$ are the actual and the predicted activities of the with molecule, respectively, and y mean is the mean of observed activity of all molecules in training set.

$$q^2 = 1 - \Sigma\left(y_i - y^{\wedge}_i\right)^2 / \Sigma\left(y_i - y_{mean}\right)^2$$

The q^2 value obtained is the indicative power of the internal stability of the model developed.

External validation

Predicted r^2 (pred_r^2) values are indicative of all the predictive powers of the developed QSAR and have been calculated using the equation (Eq A.2) where y_i and $\hat{y}_i$ are the actual and the predicted activities of the i th molecule in the test set, respectively, and y mean is the observed activity of all molecules in training set.

$$\text{Pred} - r^2 = 1 - \Sigma\left(y_i - y^{\wedge}_i\right)^2 / \Sigma\left(y_i - y_{mean}\right)^2$$

Docking studies

Molecular docking aims to predict binding modes of protein–ligand complexes, defining the preferred orientation of a molecule with respect to the others.[27] There are two primary aspects to access the quality of docking techniques: (I) docking accuracy, which perceives the genuine binding mode of the ligands to the target protein and (ii) screening advancement which measures the relative change in the identification of true binding ligands utilizing a docking strategy versus random screening.[28] Docking studies were performed on 5BR, using the GRIP docking feature in BIOPREDICTA module available in V Life MDS software 4.6. The logic of the studies lies in the fact that crystal structure of the receptor (5AR) is not available, and 5BR can serve as a substitute to the receptor as substrate for both of the enzymes is same thus, they may have same or at least similar enzymatic functions when metabolizing steroid hormones.[29] These studies would help to sort out the compounds with good binding affinity against 5AR enzyme.

Protein preparation and docking methodology

Crystal structure, with a resolution of 2.2Å, of 5BR in complex with NADPH from Homo sapiens (PDB ID: 3CAQ) was downloaded from PDB (www.rcsb.org). The protein structure was pre–processed and refined by removing the water molecules followed by addition of hydrogen atoms, deleting the ligand and co–factors present in crystal structure and finally receptor protein was taken in pdb format. GRIP docking module of Vlife MDS version 4.6. It (MDS) involves series of steps viz. pre–computation of grids, sampling a set of initial poses, search of best possible poses by maximizing favourable interaction and minimizing the steric unfavourable and repulsive interactions and was thus successfully employed to dock proposed 5–ARIs into receptor protein. The ligand forming most stable drug–receptor complex is the one with minimum dock score.

In–silico ADME prediction

Compound having the best binding interactions for target doesn't guarantee a best medication. A perfect oral medication ought to rapidly and complete absorption from the gastrointestinal tract, dispersion towards its target, metabolism in a way that does not quickly eliminate its action, and eliminate in a proper way without causing any harm.

The poor ADME has been found to be main reason for drugs failure.[30] In the present study, ADME properties were determined using pre ADMET tool version 2.0 software (preadmet.bmdrc.kr) to predict their drug like properties, where in all the analogues were neutralized and minimized before being used by software. The compounds were evaluated considering, predictive absorption for human intestinal absorption (HIA), cellular permeability (Caco–2) *in vitro*, cell permeability Maden Darby Canine Kidney (MDCK), % plasma protein binding, blood brain barrier (BBB) counting different features represented their 'druggable' pharmacokinetic profile.[31]

Results and discussion

2D–QSAR modelling and its validation (copied)

With regard to QSAR models, our main aim was to establish a predictive model with reasonable number of logical descriptors to get good generalization performance and which can be further utilized to predict the activity of present compounds and for the synthesis of steroidal molecules with potential to inhibit 5AR enzyme. Chosen molecular parameters computed for all the 29 molecules were used to create QSAR equation by relating their corresponding inhibitory activities i.e. IC_{50} values. Different models were produced used PCR, MLR and PLS and the model having best fit with least number of descriptors has considered or observed to be the best model. When this point is accomplished no further significant change in the regression coefficient (r^2 and q^2) values were observed regardless of the possibility that another descriptor is included. In the present study, PCR, MLR and PLS techniques used three, four or more factor combinations for combined dataset created around 200 equations out of which, the sensible satisfactory ones were chosen for discussion. The different models produced for better correlation by statistical analysis have been given in Table 3, though Table 4 contains different descriptors, their classes and in addition their contribution towards every descriptor. The models (1–3) were developed using simulated annealing (SA) methods, with cross correlation limit set to 0.5 and selected criteria as q^2.

Table 3 2D Statistical data of the various regression methods

Model	Simulated annealing		
	(Model 1) PCR	(Model 2) MLR	(Model 3) PLS
N	20	20	20
r^2	0.9535	0.9714	0.9669
q^2	0.8818	0.8869	0.872
F value	76.8765	73.6675	109.4339
Pred–r^2	0.5491	0.7228	0.8289

Model 1

pIC_{50}=0.2121(chi4) +0.0289 (Dipole Moment) +0.0342 (XA Most hydrophobic hydrophilic Distance) +0.0874 (SaaNE–index) –0.0634 (SsNH2count) +1.1339 (Eq B1) The generated SA–PCR model (Eq B1) showed good squared $r^2$0.9535, but low pred–$r^2$0.5491 which was not found to be satisfactory for a good correlation between the structure and activity. So, another regression method was performed in order to get improved QSAR model analysis Using same data set, a new MLR based model 2 (Eq B2) was generated.

Table 4 Contribution of various descriptors

Descriptor	Category	Contribution	Meaning
MODEL 1			
chi4	Topological	+	Signifies a retention index (fourth order) derived directly from gradient retention times.
Dipole Moment	Topological	+	Signifies dipole moment calculated from the partial charges of the molecule.
XAMostHydrophobicHydrophilicDistance	Semi–emperical	+	signifies distance between most hydrophobic and hydrophilic point on the vdW surface
SaaNE–index	Topological	#REF!	Electrotopological state indices for number of nitrogen atom connected with two aromatic bonds.
SsNH2count	Topological	#VALUE!	defines the total number of $–NH_2$ group connected with one single bond
MODEL 2			
Dipole Moment	Semi–emperical	+	Signifies dipole moment calculated from the partial charges of the molecule.
SaaNcount	Topological	0	Defines the total number of nitrogen connected with two aromatic bonds.
SdSE–index	Topological	–	Electrotopological state indices for number of sulphur atom connected with one double bond. Most hydrophilic value on the vdW surface.
SKMostHydrophilic	Semi–emperical	–	
DeltaEpsilonA	Semi–emperical	–	Signifies a retention index (fourth order) derived directly from gradient retention times.
chi4	Topological	+	
MODEL 3			
SdsNE–index	Semi–emperical	+	Electrotopological state indices for number of nitrogen atom connected with two double and one single bond.

Table Continued

Descriptor	Category	Contribution	Meaning
+vePotentialSurfaceArea	Electrostatic	–	Signifies total Vander Waals surface area with positive electrostatic potential of the molecule.
SssCH2E–index	Topological	+	Electrotopological state indices for number of –CH2 group connected with two single bonds.
SsSHcount	Electrostatic	–	Defines the total number of –SH group connected with one single bond.
YcompDipole	Semi–emperical		signifies the Y component of the dipole moment (external coordinates)
chi4	Topological		signifies a retention index (fourth order) derived directly from gradient retention times

Model 2

pIC_{50}=0.1976 (Dipole moment) +0.2108 (Saa Ncount) –1.0571 (SdSE–index) –1.8023 (SK Most hydrophilic) –4.6165(Delta epsilon) +0.4338 (chi4) –0.2958 (Eq B.2) Generated SA–MLR model 2 showed high cross–validated correlation coefficient $q^2$0.8869 between descriptors (Y comp Dipole, SAHydrophilic Area, Dipole Moment, chiV4) and 5AR inhibitory activity, but showed low pred–r^2=0.7228 and low F–value 73.6675, which was found to be considerable, but it can be improved.

Model 3

Along these lines, to enhance the nature of model, we selected another prominent regression method SA–PLS analysis. Generated SA–PLS model 3 (Eq B3) was found to be the best model and exhibits good external predictivity indicated by pred–r^2 0.8289. The squared correlation coefficient r^2 0.9669 also explains 96% of the variance in biological activity. The high F–value 109.4339 and q^2 0.8720 qualifies it to be valid model. pIC_{50}=0.0523 (Sds NE–index) –0.0046 (+ve Potential Surface Area)+0.0784 (Sss CH2E–index) +0.2945(SSSH count)–0.1044 (Y comp Dipole) +0.2972 (chi4) +1.7069 (Eq B3) The descriptors in the best model indicate effect of Sds NE–index, +ve Potential Surface Area, Sss CH2E–index, SSSH count, Y comp Dipole, chi4 on the biological activity. It is also evident from the equation that +ve Potential Surface Area is an electrostatic descriptor which signifies total vander Waals surface area with positive electrostatic potential of the molecule. It is negatively correlated with biological activity that means higher is the +ve Potential Surface Area, lower will be the inhibitory activity against 5–AR. Sds NE–Index represents the number of nitrogen atom connected with two double bond and one single bond and is positively correlated that means the higher is the number of nitrogen atom connected, the higher will be inhibitory activity against 5–AR. Chi4, a topological descriptor, is positively correlated that means the higher is the chi4, the higher will be inhibitory activity. Ycomp Dipole is a semi–empirical descriptor signifies the y component of the dipole moment (external coordinates) and the contribution of Ycomp Dipole to inhibitory activity is negative. SsSHcount represents the total number of SH groups connected with one single bond and is positively correlated with biological activity that means higher is the number of SH groups, higher will be the inhibitory activity against 5–AR. SSSCH2E–index, a topological information–based descriptor, represents number of $–CH_2$ group connected with two single bonds also contributed positively. It means more the number of $–CH_2$ groups connected with two single bonds, more will the 5AR inhibitory activity. It has also been observed that in all models, chi4 is the common descriptor and its contribution to inhibitory activity is positive. Contributions of the various descriptors, fitness plots and actual and predicted activity of the training and sets molecules w.r.t. SA–PLS model are shown in Figure 1 & 2 and Figure 3A & 3B.

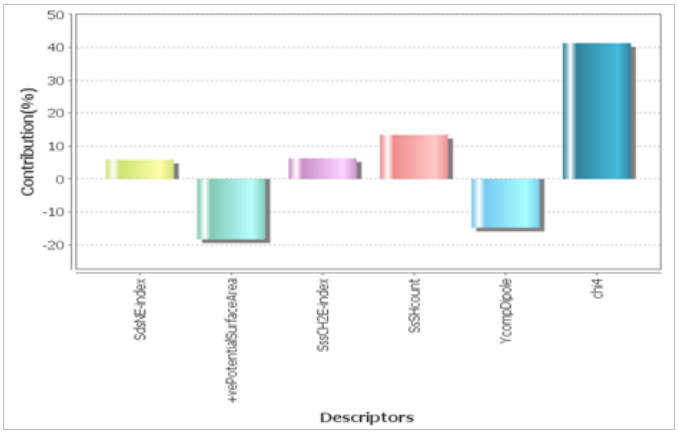


Figure 3 Contribution plot of various descriptors in SA–PLS.

3D–QSAR

The kNN system depends on a simple distance learning approach whereby an unknown member is classified by the majority of its k–nearest neighbours in the training set, which is measured by a suitable distance metric (e.g., a molecular similarity measure calculated using field interactions of molecular structures).[32,33] In the present study, kNN–MFA model was developed coupled with SA method to develop 3D–QSAR models of selected 29 molecules as 5ARIs based on steric, hydrophobic and electrostatic fields. The best kNN–MFA 3D–QSAR model exhibited good cross–validated correlation coefficient q^2= 0.9523 and external predictivity pred–r^2= 0.8086 Table 5. The fitness plot and plots of observed versus predicted activity of both training and test sets molecules helped in cross–validation of kNN–MFA QSAR model were depicted in Figure 4, Figure 5A & 5B**.** In 3D–QSAR studies, 3D data points were generated and pharmacophore were used to optimize the electrostatic and steric requirements for the 5AR inhibitory activity. The ranges of property values were based on the variation of the field values at the chosen points using the most active molecule and its nearest neighbour set. The presence of blue balls represents electrostatic parameter and its negative range indicates that negative electrostatic potential is favourable for increase in the activity and hence a relatively more electronegative substitution is preferred in that region and vice–versa. Similarly green balls are presenting steric groups and its negative range indicates that negative steric potential is

favourable for increase in the activity and hence a relatively less bulky substituent group is preferred in that region. Yellow balls indicate the hydrophobic groups. 3D–QSAR kNN–MFA based model of selected series of 29 molecules has shown 5 descriptors (3 electronic, 1 steric and 1 hydrophobic) contributing to electrostatic, hydrophobic and steric parameter namely E_412, E_399, E_349, H_1100, and S_397 respectively. Here, electronic descriptors have been found to play major role as compared to steric descriptor. The electronic descriptor have been contributed by presence of electron withdrawing groups at position 17 of steroidal nucleus and which is further in agreement with their negative values as shown in Figure 6A & 6B, whereas the steric descriptor has been contributed by lesser bulkier groups at position 3 of steroidal nucleus. The lattice point E_412 and steric S_397 are of great interest and suggested to synthesize novel 5ARI having electron withdrawing groups at C–17.

Table 5 3D statistical analysis of kNN regression method

3D kNN–MFA method	
k Nearest Neighbour	2
N	20
Degree of freedom	14
q^2	0.9523
Pred–r^2	0.8068

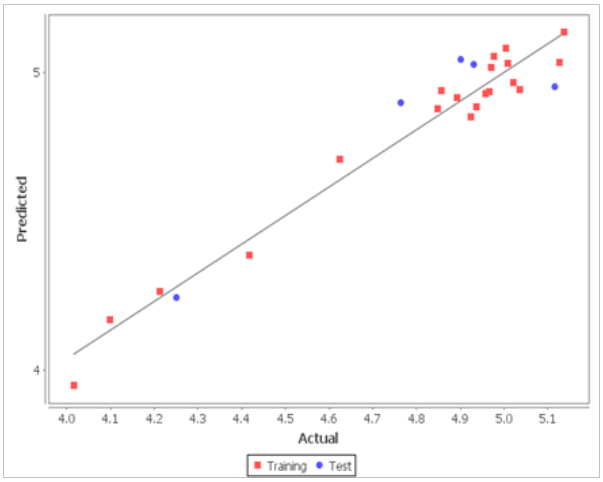
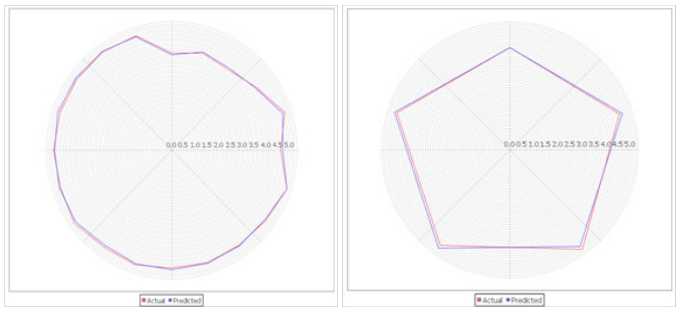


Figure 4 Fitness plot of training and test set molecules by PLS method.



(A) **(B)**

Figure 5(A–B) Radar plot showing overlapping of actual activity of training and test set molecules by PLS method. Radar plot showing overlapping of predicted activity of training and test set molecules by PLS method.

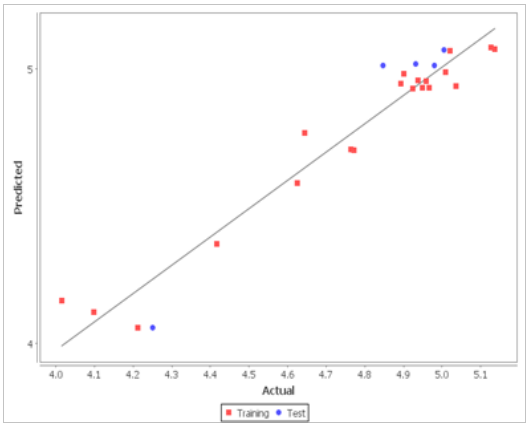


Figure 6 Fitness plot of training and test set molecules by kNN–MFA method.

Table 6 Docking score of the ligands

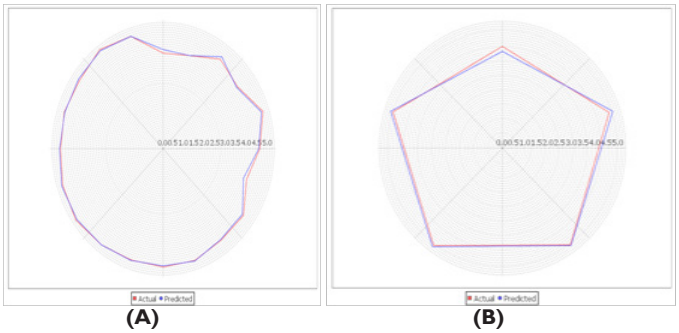
Compound	D–Score
1a	–76.988
1b	–88.618
1c	–80.937
1d	–81.907
1e	–72.676
2a	–69.776
2b	–74.616
2c	–76.344
2d	–56.346
2e	–85.26
3	–74.728
4a	–57.153
4b	–56.199
5a	–62.894
5b	–64.128
5c	–58.282
5d	–67.534
5e	–66.34
5f	–58.192
5g	–63.047
5h	–61.251
6a	–74.742
6b	–53.456
6c	–58.193
6d	–62.499
6e	–59.663
6f	–61.299
6g	–66.938
6h	–56.274

Molecular docking

Docking, the prediction of binding mode of small molecule ligands to protein targets is central to numerous biological processes and is of fundamental importance in modern structure–based drug design (SBDD). Docking simulations (*in–silico*) studies were performed on the 29 compounds using extra precision GRIP docking feature in BIOPREDICTA module available in V Life MDS software package, version 4.6 in order to postulate a hypothetical binding model for their interaction with active site of 5BR receptors keeping external ligand as reference for comparison. Table 6 shows the docking scores obtained for all 29 molecules. Compounds 1b, 3 and 5g are selected on the basis of low IC_{50} values and their 2D and 3D representation of the ligand interaction w.r.t. 5BR receptor respectively are shown in Figure 7(A–C) & Figure 8(A–C). Grip values docked pose of the fitted ligands were visualized extending deep into the active site pocket and showing several interactions such as Vander Waal's, hydrophobic contacts and π–π stacking interactions and hydrogen bonds with the key residues of the active site. All the compounds have been found to bound best with the 5BR receptor affording highest dock score from –88.618 to –53.456. The test ligands showed similar docking behaviour with reference external ligand majorly with by interacting with common amino acid residues Lys 273A, Thr 224A, Ser 220A, Leu 222A strong hydrophobic bonds. Various hydrophobic and hydrogen interaction of test ligand with the protein is shown in the Table 7. Compound 1b has shown greater affinity for 5BR receptor with highest dock score of –88.618. Compound **1b** has showed hydrophobic and hydrogen interactions, with protein (3CAQ) such as SER 225A, LYS 273A, TYR 219A, TYR 26A, PRO 221A, SER 220A, ARG279A, SER274A, ALA59A, and THR 224A. The high D–scores of the ligands with respective to reference can not only be attributed to their complementarily shape, but 60–70% similarity of residues with reference could be responsible for their great binding.

Table 7 Various hydrophobic and hydrogen interaction of test ligands

Ligand pose	D–Score	No. of residues	Hydrophobic	Hydrogen
Compound 1b	–88.618	10	ARG 279A, SER 274A, SER 225A, LYS 273A, THR 224A, TYR 219A, TYR 26A, PRO 221A, SER 220A, LEU 222A	–
Compound 3	–74.728	13	LEU 222A, ILE 271A, GLY 223A, LYS 273A, PRO 272A, PRO 221A, TYR 21A, GLY 24A, SER 225A, SER 220A, THR 224A, TYR 219A	ASP 53A
Compound 5g	–63.047	8	LYS 273A, THR 224A, LEU 239A, LEU 222A, ARG 279A, SER 225A	GLU 28A, SER 220A, ARG 279A, LYS 273A



(A) **(B)**

Figure 7(A–B) Radar plot showing overlapping of actual and predicted activity of training set molecules by kNN– MFA model. Radar plot showing overlapping of actual and predicted activity of test set molecules by kNN–MFA model**.**

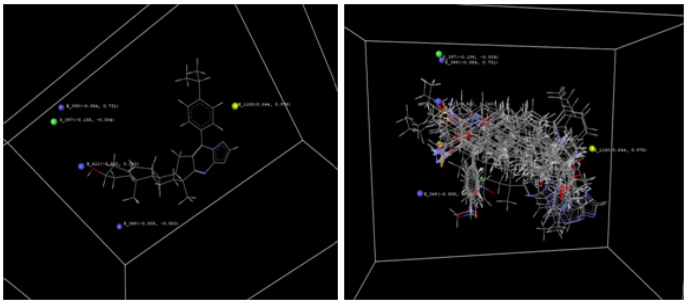
In–silico ADME prediction

Number of the drugs under clinical trials couldn't see the clinics because of failure at the phase of pharmacokinetic assessment. Initial screening of hits and leads before their clinical testing will not only decrease the rate of failure, but it reduces the cost of drugs discovery program. Taking into consideration, a preliminary predictive *in–silico* pharmacokinetic study of the selected series of compounds was undertaken using online server pre ADMET. Incorporation of such tools as a part of the drug design process can screen molecules that are more likely to exhibit satisfactory ADME properties. The server calculated the parameters such as HIA (%), Caco–2 (nm/sec), MDCK (nm/sec), plasma protein binding (%), BBB (log PS). For ADME prediction, 3 molecules (compound 1b, 3 & 5g) from the selected series of 29 molecules were chosen on the basis of low IC_{50} values and results are expressed in Table 8**.** These properties are determinant for drug development factors (ADME), mainly, human intestinal absorption properties, because it is determinant for the drug development that purport to be administered orally. All the compounds under study presented human intestinal absorption value (HIA) in the range of 93.3093 to 98.6105. The absorption process is further related to the permeation of compounds through biological membrane under the influence of physicochemical characteristics, and present observations indicate the good physicochemical properties of the compounds and enabled them to qualified HIA% with values >80–100%. Also the low values for brain/blood partition coefficient were found indicating that they will have a very low potential to cross the brain/blood brain barrier there by eliminating the possibility of CNS related toxicity. Further, the cell permeability *in vitro* Caco–2 is an important test to assess intestinal absorption of drugs. The results indicate the cell permeability of all the selected compounds ranges from 21.2951 to 32.2115. Further, analysing the data in this

MDCK system of the various inhibitors, present obtained indicates that all the compounds have low permeability towards kidney cells with values ranging from 0.04344nm/sec to 6.60536nm/sec. The distribution properties of plasma protein binding (PPB %) verified that the all compounds showed binding to plasma proteins i.e. (ranges from 93.10095 to 100%) and such inhibitors would have a large amount of free drug to interact with their receptors, triggering the form of this pharmacological response. In addition to changing the pharmacological response of molecule the PPB also modifies the renal excretion because only unbound drug is available for glomerular filtration thus increasing excretion and decreasing half–life.

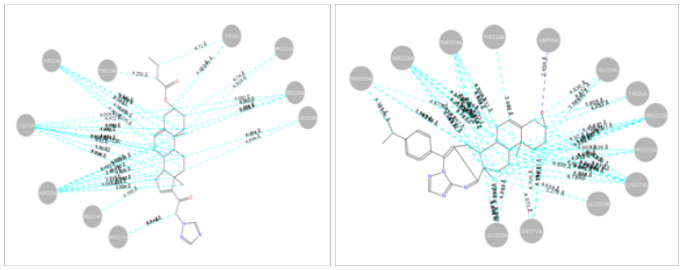
Table 8 Pre–ADME prediction of ligands

Compound	HIA (%)	Caco–2 (Nm/Sec)	MDCK (Nm/Sec)	BBB (Log PS)	Plasma protein Binding (%)
1b	98.6105	30.8223	0.05505	0.0305	93.101
3	96.9445	32.2115	6.60536	2.9061	96.739
5g	93.3093	21.2951	0.04344	0.3076	100

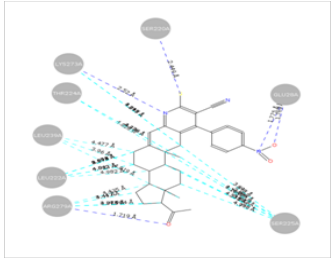


(A) (B)

Figure 8(A–B) Contribution plot of steric and electrostatic field of interactions 3D view of aligned molecules. 3D view of active molecule (3D–QSAR) by kNN–MFA model.

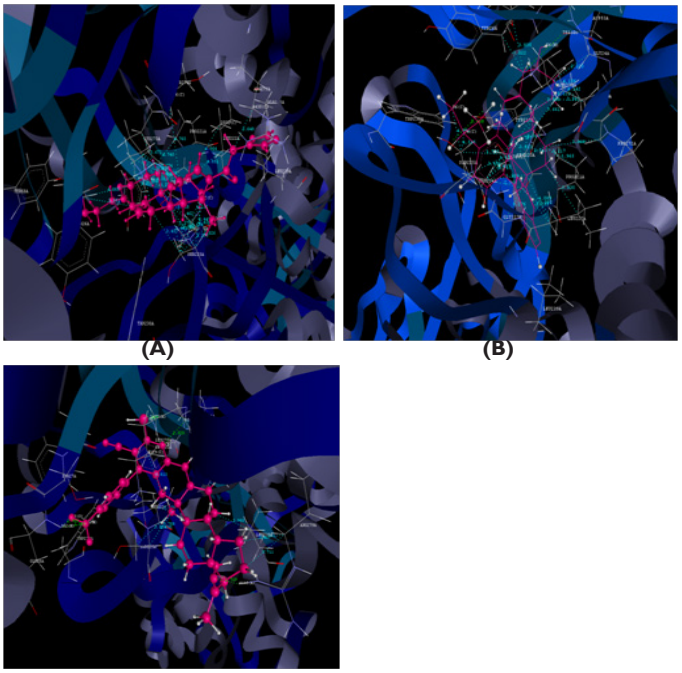


(A) (B)



(C)

Figure 9(A–C) 2D representation showing the interaction of the compounds 1b, 3 and 5g with active amino residues of 5β–reductase receptor respectively.



(A) (B) (C)

Figure 10(A–C) 3D representation showing the interaction of the compounds 1b, 3 and 5gwith active amino residues of 5β–reductase receptor respectively.

Conclusion

Despite the lack of structural information on 5–AR, the design of potent inhibitors can be attempted by means of well–established QSAR techniques. In the present study chemical features of a series of compounds along with their activities ranging over several orders of magnitudes was used to generate QSAR models, and these could be further utilized to predict the activity of new designed compounds. 3D–QSAR kNN based model with good statistical data having q^2 approximately 95% (internal validation) and 80% (external validation) has demonstrated the importance of electronic and less steric feature at position 17 and 3 of the steroid nucleus respectively. Further docking analysis using D–score and ligand receptor interactions showed that all the studied compounds are well accommodated in the binding pocket of 5BR and variations in the activity are dominated by hydrogen and

hydrophobic interactions. Compound 1b bind and interact with lowest estimated binding energy –88.16 with its hydrogen bonding and hydrophobic interactions and the observations were further found in agreement with its good ADME properties. This study suggested that the strategy of introducing important physicochemical features and orientation for developing novel potent 5AR inhibitors.

Acknowledgments

The authors are thankful to the Dr Amit Bedi, and Dr. Kundan; Vlife Science Technologies Pvt. Ltd., Pune, India, for providing us with the VlifeMDS4.6 software and for their continual support in the study.

Conflict of interest

Author declares that there is no conflict of interest.

References

1. Payne AH, Hales DB. Overview of steroidogenic enzymes in the pathway from cholesterol to active steroid hormones. *Endocrine reviews*. 2004;25(6):947–970.
2. Ibrahim–Ouali M. First total synthesis of 11–selena steroids. *Tetrahedron Letters*. 2009;50(14):1607–1609.
3. Dubey RK, Oparil S, Imthurn B, et al. Sex hormones and hypertension. *Cardiovascular research*. 2002;53(3):688–708.
4. Latham KA, Zamora A, Drought H, et al. Estradiol treatment redirects the isotype of the autoantibody response and prevents the development of autoimmune arthritis. *The Journal of Immunology*. 2003;171(11):5820–5827.
5. Sheridan PJ, Blum K, Trachtenberg M. Steroid receptors and disease. *Plenum Press*. 1988:289–564.
6. Moudgil VK. Steroid receptors in health and disease. *Plenum Press*. 1987.
7. Malley BW. Hormones and signalling. *Academic Press*. 1998.
8. Parker MG. Steroid hormone action. *Irl Press*. 1993.
9. Tiwari A, Krishna NS, Nanda K, et al. Benign prostatic hyperplasia: an insight into current investigational medical therapies. *Expert opinion on investigational drugs*. 2005;14(11):1359–1372.
10. Benign prostatic hyperplasia. *Nature Reviews Disease Primers*. 2016;(2):16032.
11. Aggarwal S, Thareja S, Verma A, et al. An overview on 5α–reductase inhibitors. *Steroids*. 2010;75(2):109–153.
12. Li J, Ding Z, Wang Z, et al. Androgen regulation of 5α–reductase isoenzymes in prostate cancer: implications for prostate cancer prevention. *PLoS One*. 2011;6(12):e28840.
13. Uemura M, Tamura K, Chung S, et al. Novel 5α-steroid reductase (SRD5A3, type-3) is over expressed in hormone-refractory prostate cancer. *Cancer science*. 2008;99(1):81–86.
14. Liang T, Cascieri MA, Cheung AH, et al. Species differences in prostatic steroid 5α–reductases of rat, dog, and human. *Endocrinology*. 1985;117(2):571–579.
15. Bratoeff E, Garrido M, Ramírez–Apan T, et al. Effect of dehydroepiandrosterone derivatives on the activity of 5α–reductase isoenzymes and on cancer cell line PC–3. *Bioorganic & medicinal chemistry*. 2014;22(21):6233–6241.
16. Bratoeff E, García P, Heuze Y, et al. Molecular interactions of progesterone derivatives with 5α–reductase types 1 and 2 and androgen receptors. *Steroids*. 2010;75(7):499–505.
17. Fitzpatrick JM, Artibani W. Therapeutic strategies for managing BPH progression. *European urology supplements*. 2006;5(20):997–1003.
18. Hansch C, Sammes PG, Taylor JB, et al. Comprehensive medicinal chemistry, Quantitative Drug Design. *Oxford*. 1990;272:250–202.
19. Vedani A, Descloux AV, Spreafico M, et al. Predicting the toxic potential of drugs and chemicals *in silico*: A model for the peroxisome proliferator–activated receptor γ (PPAR γ). *Toxicology letters*. 2007;173(1):17–23.
20. Silva–Ortiz AV, Bratoeff E, Ramírez–Apan T, et al. Synthesis and activity of novel 16–dehydropregnenolone acetate derivatives as inhibitors of type 1 5α–reductase and on cancer cell line SK–LU–1. *Bioorganic & medicinal chemistry*. 2015;23(24):7535–7542.
21. Silva–Ortiz AV, Bratoeff E, Ramírez–Apan T, et al. Synthesis of new derivatives of 21–imidazolyl–16–dehydropregnenolone as inhibitors of 5α–reductase 2 and with cytotoxic activity in cancer cells. *Bioorganic & medicinal chemistry*. 2017;25(5):1600–1607.
22. Tantawy MA, Nafie MS, Elmegeed GA, et al. Auspicious role of the steroidal heterocyclic derivatives as a platform for anti–cancer drugs. *Bioorganic chemistry*. 2017;73:128–146.
23. Amr AE, Abdalla MM, Hussein MM, et al. 5α–reductase inhibitors and anti–prostate cancer activities of some synthesized 4–(aryl)–4–pregneno [3, 2–e] pyridinone derivatives. *Biomedical Research*. 2017;28(2):601–607.
24. Briceland LL, Pasko MT, Mylotte JM. Serum bactericidal rate as measure of antibiotic interactions. *Antimicrobial agents and chemotherapy*. 1987;31(5):679–685.
25. Shen M, LeTiran A, Xiao Y, et al. Quantitative structure−activity relationship analysis of functionalized amino acid anticonvulsant agents using k nearest neighbor and simulated annealing PLS methods. *Journal of medicinal chemistry*. 2002;45(13):2811–2823.
26. Metropolis N, Rosenbluth AW, Rosenbluth MN, et al. Equation of state calculations by fast computing machines. *The journal of chemical physics*. 1953;21(6):1087–1092.
27. Meng XY, Zhang HX, Mezei M, et al. Molecular docking: a powerful approach for structure–based drug discovery. *Current computer–aided drug design*. 2011;7(2):146–157.
28. Muthas D, Sabnis YA, Lundborg M, et al. Is it possible to increase hit rates in structure–based virtual screening by pharmacophore filtering? An investigation of the advantages and pitfalls of post–filtering. *Journal of Molecular Graphics and Modelling*. 2008;26(8):1237–1251.
29. Yao Z, Xu Y, Zhang M, et al. Discovery of a novel hybrid from finasteride and episteride as 5α–reductase inhibitor. *Bioorganic & medicinal chemistry letters*. 2011;21(1):475–478.
30. Hodgson J. ADMET—turning chemicals into drugs. *Nature Biotechnology*. 2001;19(8):722.
31. Silva NS, Gonçalves LK, Duarte JL, et al. Computational analysis of physicochemical, pharmacokinetic and toxicological properties of deoxyhypusine synthase inhibitors with antimalarial activity. *Computational Molecular Bioscience*. 2014;4(4):47–57.
32. Sharaf MA, Illman DL, Kowalski BR. Chemometrics. *John Wiley & Sons*. 1986:1–352.
33. Sun LX, Xie YL, Song XH, et al. Cluster analysis by simulated annealing. *Computers & Chemistry*. 1994;18(2):103–108.