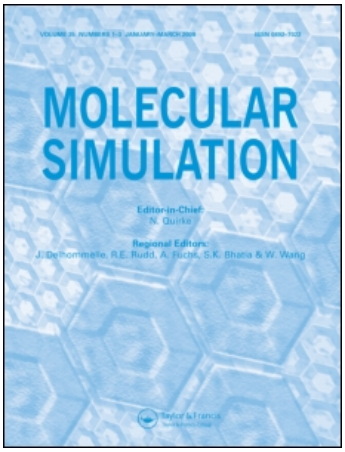

Molecular Simulation
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713644482

Classical and 3D-QSAR studies of cytochrome 17 inhibitor imidazole-substituted biphenyls

Partha Pratim Roy[a]; Kunal Roy[a]
[a] Division of Medicinal and Pharmaceutical Chemistry, Drug Theoretics and Cheminformatics Lab, Department of Pharmaceutical Technology, Jadavpur University, Kolkata, India

First published on: 23 November 2009

To cite this Article Roy, Partha Pratim and Roy, Kunal(2010) 'Classical and 3D-QSAR studies of cytochrome 17 inhibitor imidazole-substituted biphenyls', Molecular Simulation, 36: 4, 311 — 325, First published on: 23 November 2009 (iFirst)
To link to this Article: DOI: 10.1080/08927020903426493
URL: http://dx.doi.org/10.1080/08927020903426493

Molecular Simulation
Vol. 36, No. 4, April 2010, 311–325

Taylor & Francis
Taylor & Francis Group

Classical and 3D-QSAR studies of cytochrome 17 inhibitor imidazole-substituted biphenyls

Partha Pratim Roy and Kunal Roy*

Division of Medicinal and Pharmaceutical Chemistry, Drug Theoretics and Cheminformatics Lab, Department of Pharmaceutical Technology, Jadavpur University, Kolkata 700 032, India

(*Received 24 August 2009; final version received 16 October 2009*)

Classical and 3D-quantitative structure–activity relationship (3D-QSAR) studies have been performed for a series of imidazole-substituted biphenyl derivatives ($n = 28$) having 17α-hydroxylase/17-20-lyase (CYP17) inhibitory activity. 3D-QSAR analyses were performed using molecular shape analysis (MSA), receptor surface analysis (RSA) and molecular field analysis (MFA). The chemometric tools used for the analysis are genetic function approximation (classical QSAR and MSA) and genetic partial least squares (MFA and RSA). The developed QSAR models suggest that the presence of an imidazole group at the *para* position of the biphenyl ring is favourable for the inhibitory activity. The MSA-derived model indicates the importance of different shadow parameters (Lx, Syz) as well as molecular rigidity. The MFA model suggests the interaction points near the phenyl ring and the imidazole ring, while the RSA model indicates the nature of the hypothetical receptor surface around the ring systems. The RSA-derived model was found to be the best model based on the highest internal ($Q^2 = 0.932$) predictive power as well as $r^2_{m(\text{overall})}$ criteria (0.858). According to external prediction statistics ($R^2_{\text{pred}} = 0.876$), the MFA-derived model outperforms other models. Overall, there should be small (—H, —OH, —NH_2) hydrophilic hydrogen bond donor groups at the terminal phenyl ring (*meta* or *para* position) for optimum inhibitory activity.

Keywords: QSAR; LFER; MSA; MFA; RSA

Introduction

Genetic susceptibility in the progression of cancer is an important area to be explored, especially in hormone-related carcinomas such as breast and prostate cancer (PC) [1]. PC is currently the most common malignancy and age-related cause of death in elderly men worldwide [2]. Regardless of advancements in early detection and treatment methods, about 232,090 new cases and nearly 30,350 deaths occurred in the United States in 2005 alone [3]. According to the American Chemical Society, during 2009, about 192,280 new cases of PC will be diagnosed and about 27,360 men will die in the USA (http://www.cancer.org/docroot/CRI/content/CRI_2_4_1X_What_are_the_key_statistics_for_prostate_cancer_36.asp). In 1941, Charles Huggins and colleagues [4,5] first demonstrated the benefits of hormone deprivation therapy for PC. Androgen plays a crucial role in the growth development of the prostate gland, and approximately 80% of human prostate tumours are androgen dependent. Androgen ablation remains the standard treatment for PC. The presently used treatment is surgical castration (orchidectomy) or, its medical equivalent, the application of gonadotropin-releasing hormone (GnRH) analogues to suppress testicular androgen biosynthesis [6]. But, unfortunately, GnRH agonists fail to prevent the synthesis of testosterone by adrenal glands. Then, the concept of 'combined androgen blockade', i.e. the combination of GnRH analogues with androgen receptor (AR) antagonists (flutamide, cyproterone acetate) to reduce the stimulatory effects of remaining androgens, was often practised [7,8]. But anti-androgens can act as weak agonists in PC cells expressing mutated and/or overexpressed AR [9]. In addition, combination therapy with anti-androgen seems to fail to extend survival rates in patients with advanced PC [10].

Cytochrome P450s (CYPs) represent a large class of haem-containing enzymes that catalyse the metabolism of multitudes of substrates, both endogenous and exogenous. The majority of CYPs (designated as classes I and II) act as versatile monooxygenases. Microsomal CYP enzymes catalyse specific steps in the biosynthesis of steroid hormones, cholesterol, prostanoids and bile acids, participate in the catabolism of endogenous compounds, including fatty acids and steroids, and are involved in the degradation of exogenous compounds, including a wide variety of structurally diverse drugs and carcinogens [11–15].

Most cytochromes (CYPs) were once considered to be liver-specific enzymes, but now their extrahepatic expression has been well established. Members of the CYP1–3 families have been identified in both healthy and cancerous extrahepatic tissues [16]. 17α-Hydroxylase/17-20-lyase (CYP17) is responsible for the production of androgen expressed in testes and adrenals [17]. The cytochrome b_5-modulated key enzyme CYP17 for

*Corresponding author. Email: kunalroy_in@yahoo.com

ISSN 0892-7022 print/ISSN 1029-0435 online

DOI: 10.1080/08927020903426493
http://www.informaworld.com

androgen synthesis catalyses the last two sequential reactions in the production of testosterone such as (1) 17α-hydroxylation and (2) subsequent C17–C20 bond cleavage of pregnenolone and progesterone to form dihydroandrostendione (DHEA) and androstenedione, the precursors of testosterone [18,19]. Therefore, CYP17 has become the target of interest for the systemic inhibition of androgen production and CYP17 inhibitors could be a promising alternative to GnRH analogues and anti-androgens for the treatment of PC.

Ketoconazole, an antifungal agent, was the only CYP17 inhibitor used clinically for the treatment of advanced PC to reduce testosterone biosynthesis through unspecific inhibition [20,21]. Different attempts were made in order to obtain specific steroidal as well as non-steroidal CYP17 inhibitors. Recently, the steroidal CYP17 inhibitor abiraterone has entered into phase II clinical trial showing high activity in postdocetaxel castration refractory PC patients with no dose-limiting toxicity [22,23]. Being a potent inhibitor of CYP17, steroidal compounds showed certain serious limitations such as short half-life, poor bioavailability and first-pass metabolism during oral administration [24–29]. Because of the adverse effects related to the steroidal nucleus, there is a need to develop new potent and selective non-steroidal CYP17 inhibitors. Several categories of highly potent and specific steroidal and non-steroidal CYP17 inhibitors have been reported in the literature [23,30,31]. The molecular scaffold for both steroidal and non-steroidal inhibitors is an iron-complexing group, mostly a nitrogen-bearing heterocycle, combined with a lipophilic moiety with an appropriate hydrophilic substitution at an appropriate position [32–37].

The design of selective and potent inhibitors of CYP17 for the treatment of PC is of utmost importance in accelerating drug discovery process. Due to the non-availability of the crystal structure of the CYP17 enzyme until now, only homology modelling [using CYP X-ray (3DBG and 1N97) from bacteria] [38,39] followed by docking studies has been performed in order to identify the interacting amino acids in the active site [23,30,31]. In the present work, we perform the molecular shape analysis (MSA), the molecular field analysis (MFA) as well as the receptor surface analysis (RSA) of a series of imidazole-substituted biphenyls reported by Wachall et al. [40], in order to explore the contribution of the molecular shape as well as other physicochemical features of the ligands to the human CYP17 inhibitory activity and the essential information of the hypothetical active site of the human CYP17 enzyme. In addition, we also perform the quantitative structure–activity relationship (QSAR) analysis using a classical linear free energy-related model of Hansch to design potent and selective CYP17 inhibitors. We have not clubbed the present data-set [40] with other data-sets having CYP17 inhibitory activity reported in the literature due to differences in the experimental protocols.

Material and methods

Data-set and descriptors

Inhibitory activities of a series of imidazole-substituted biphenyl derivatives (Figure 1; Table 1) against the human CYP17 testicular enzyme reported by Wachall et al. [40] were used as the model data-set for the present QSAR analysis. The inhibitory potencies of the compounds [IC_{50} (μM)] were converted to the logarithmic scale [pIC_{50} (mM)] and then used for subsequent QSAR analyses as the response variable. For maintaining data homogeneity, we have not clubbed this data-set with those reported elsewhere in the literature as the protocols for the experiments were found to be different. For the classical QSAR analysis using a linear free energy-related (LFER) model of Hansch [41,42], we took lipophilicity constant (π), electronic constant (Hammett σ) and steric [molar refractivity (MR) and STERIMOL L, B_1 and B_5] parameters of the aryl ring substituents along with appropriate indicator parameters. Different physicochemical and indicator variables used in this study are defined in Table 2. The values of the physiochemical substituent constants (Table 3) were taken from the literature [43]. The MSA, RSA and MFA were used as the 3D-QSAR techniques. For the MSA, shape (DiFFV, Fo, NCOSV, COSV, ShapeRMS), electronic (Apol, Dipole, HOMO, LUMO and Sr) and spatial (Radius of gyration, Jurs, Shadow indices, Area, PMI-mag, Density, Vm) descriptors were used. For the calculation of 3D descriptors, multiple conformations of each molecule were generated using the optimal search as the conformational search method. Each conformer was subjected to an energy minimisation procedure using a smart minimiser under an open force field (OFF) to generate the lowest energy conformation for each structure. The charges were calculated according to the Gasteiger method. All the descriptors were calculated using Descriptor+module of the Cerius2 version 4.10 software [44] under QSAR+environment on a Silicon Graphics O2 workstation running under the IRIX 6.5 operating system. Definitions of all the descriptors can be found at the Cerius2 tutorial available at the website http://www.accelrys.com.

Model development

The data-set (n=28) was first divided into two subsets by a K-means clustering technique based on a standardised physicochemical descriptor matrix [45]. The number of compounds in the training set was 21, while the number of compounds in the test set was 7. In the case of the Hansch

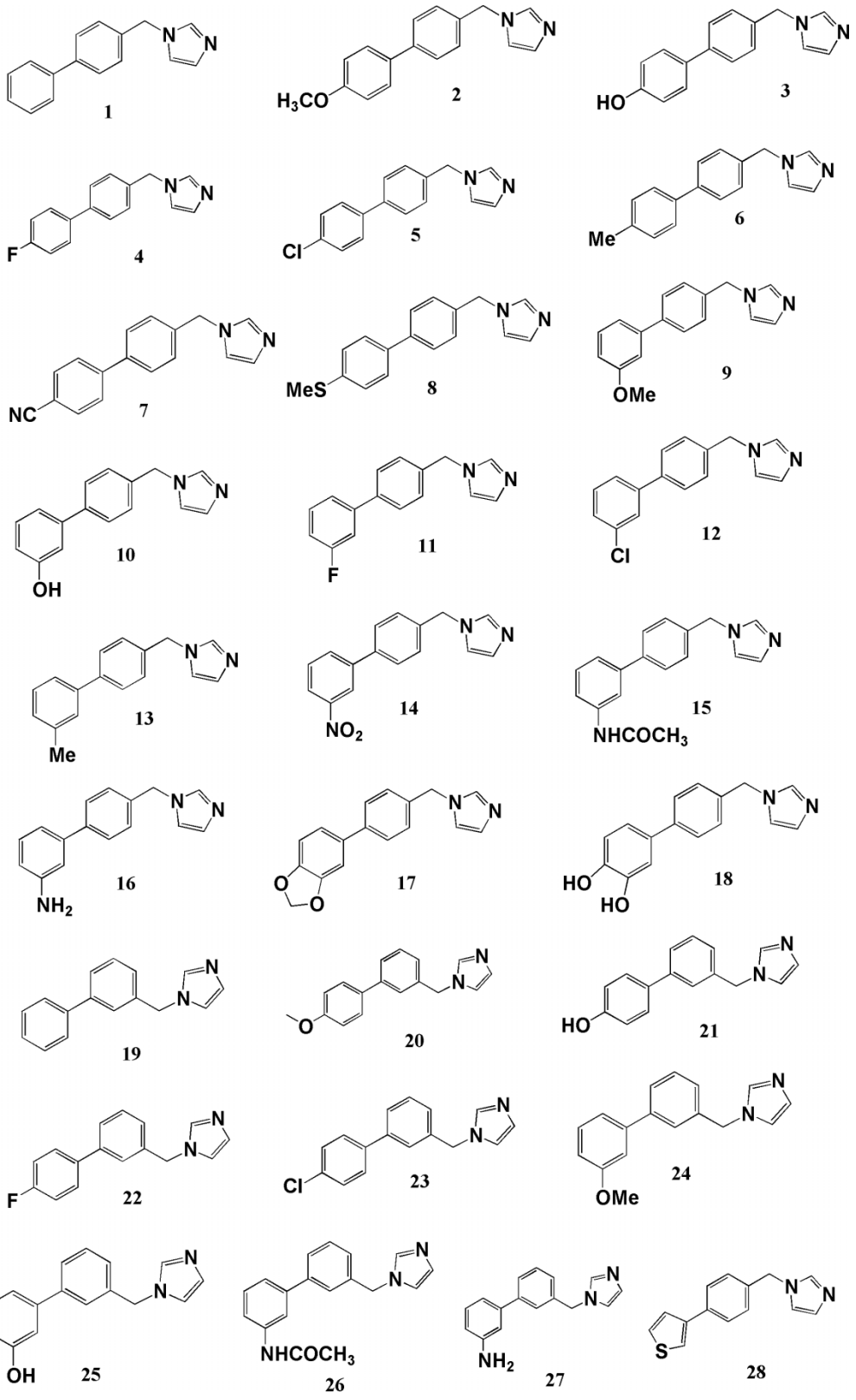

Figure 1. Molecular structures of imidazole-substituted biphenyls (**1**–**28**).

Table 1. Observed and calculated CYP17 inhibitory activity.

Sl	Obs[a]	Cal[b]	Cal[c]	Cal[d]	Cal[e]
Training set					
01	3.009	3.103	2.982	3.101	3.049
02	2.432	2.487	2.308	2.257	2.593
03	3.509	3.541	3.462	3.590	3.574
05	2.237	2.243	2.336	2.423	2.206
06	2.377	2.670	2.444	2.436	2.302
07	2.602	2.721	2.444	2.577	2.586
08	2.114	1.913	2.202	2.102	2.278
09	3.229	3.111	3.128	3.169	3.214
10	3.886	3.384	3.733	3.592	3.946
11	3.18	3.044	3.000	3.059	3.158
12	2.886	2.805	2.976	2.914	2.812
14	2.745	3.220	2.629	2.796	2.593
16	3.678	3.619	3.559	3.729	3.663
17	2.509	2.528	2.586	2.737	2.793
19	2.658	2.541	2.776	2.605	2.593
21	3.066	2.980	3.285	3.166	3.002
22	2.538	2.450	2.390	2.539	2.470
24	2.432	2.550	2.597	2.288	2.479
25	2.921	2.823	3.248	2.711	2.850
26	2.678	2.949	2.570	2.987	2.666
28	3.481	–	3.513	3.386	3.339
Test set					
04	3.018	3.011	2.698	3.023	2.932
13	2.886	2.868	2.783	2.891	3.086
15	3.638	3.510	3.412	3.789	3.355
18	4.06	3.261	4.010	3.674	4.183
20	2.347	1.926	2.173	2.306	2.593
23	1.886	1.681	1.964	2.028	2.168
27	2.377	3.058	3.064	2.868	2.862

[a] Obs, observed CYP17 inhibitory activity [40]. [b] Cal, calculated from Equation (1). [c] Cal, calculated from Equation (2). [d] Cal, calculated from Equation (3). [e] Cal, calculated from Equation (4).

method, one compound in the training set could not be considered due to the absence of a common scaffold. For the development of the QSAR models, the statistical techniques used were genetic function approximation (GFA) and genetic partial least squares (G/PLS).

Table 2. Definitions of physicochemical and indicator parameters.

Parameter	Definition
π_m	π value of the substituent at the *meta* position of the terminal phenyl ring
σ_m	Hammett σ constant value of the substituent at the *meta* position of the terminal phenyl ring
MR_m	MR value of the substituent at the *meta* position of the terminal phenyl ring
π_p	π value of the substituent at the *para* position of the terminal phenyl ring
σ_p	Hammett σ constant value of the substituent at the *para* position of the terminal phenyl ring
MR_p	MR value of the substituent at the *para* position of the terminal phenyl ring
L_p	L_p is the length of the substitutions at the *para* position of the terminal phenyl ring
$I_{Imidazole}$	Indicator variable having a value of 1 for the *para* substitution of the imidazole group and 0 for the *meta* substitution

For the MSA, the major steps are as follows: (1) generating conformers and energy minimisation; (2) hypothesising an active conformer (global minimum of the most active compound); (3) selecting a candidate shape reference compound (based on active conformation); (4) performing pairwise molecular superimposition using the maximum common subgroup (MCSG) method; (5) measuring molecular shape commonality using MSA descriptors; (6) determining other molecular features by calculating spatial, electronic and conformational parameters; (7) selecting conformers and (8) generating QSAR equations by the GFA. Optimal search was used as the conformational search method. Each conformer was subjected to an energy minimisation procedure using a smart minimiser under an OFF to generate the lowest energy conformation for each structure. The global minimum conformer of the most active compound (**10**) was selected as a shape reference to

Table 3. Values of physicochemical parameters (substituent constants)[a].

Substituents	π	σ_m	σ_p	MR[b]	B_1	B_5	L
—OMe	−0.02	0.12	−0.27	0.787	1.35	3.07	3.98
—OH	−0.67	0.12	−0.37	0.285	1.35	1.93	2.74
—F	0.14	0.34	0.06	0.092	1.35	1.35	2.65
—Cl	0.71	0.37	0.23	0.603	1.80	1.80	3.52
—Me	0.56	−0.07	−0.17	0.565	1.52	2.04	2.87
—CN	−0.57	0.56	0.66	0.633	1.60	1.60	4.23
—SMe	0.61	0.15	0.00	1.382	1.70	3.26	4.30
—NO_2	−0.28	0.71	0.78	0.736	1.70	2.44	3.44
—$NHCOCH_3$	−0.97	0.21	0.00	1.493	1.35	3.61	5.09
—NH_2	−1.23	−0.16	−0.66	0.542	1.35	1.97	2.78
—OCH_2O—	−0.05	−0.16	−0.16	0.90	1.35	3.07	3.98
H	0.00	0.00	0.00	0.10	1.00	1.00	2.06

[a] Taken from [41]. [b] MR values scaled to a factor of 0.1 as usual.

which all the structures in the study compounds were aligned through pairwise superpositioning. The method used for performing the alignment was MCSG [44,46]. The measure of molecular shape similarity is interdependent upon the molecular superposition. Finally, additional electronic, spatial and thermodynamic descriptors were also calculated.

MFA is a method for quantifying the interaction energy between a probe and a set of aligned molecules [47]. The major steps of the MFA [46] are as follows: (1) generating conformers and energy minimisation; (2) matching atoms using the MCSG alignment; (3) setting MFA preferences (rectangular grid with a 2 Å step size, charges by the Gasteiger algorithm, H^+, CH_3, CH_3^+, CH_3^- and donor/acceptor as probes); (4) creating the field and (5) analysing by the G/PLS method. MFA models are predictive and sufficiently reliable to guide the chemist in the design of novel compounds. This approach is effective for the analysis of data-sets where activity information is available but the structure of the receptor site is unknown. MFA attempts to postulate and represent the essential features of a receptor site from the aligned common features of the molecules that bind to it. This method generates multiple models that can be checked easily for validity. The MFA formalism calculates probe interaction energies on a rectangular grid around a bundle of active molecules. Atoms in the target molecule are fixed, so that intramolecular energy in the target is ignored. The surface is generated from a 'Shape Field'. The atomic coordinates of the contributing models are used to compute field values on each point of a 3D grid. Fields of molecules are represented using grids in the MFA, and each energy associated with an MFA grid point can serve as the input for the calculation of a QSAR. These energies were added to the study table to form new columns headed according to the probe type. For a set of structures for which energy fields are generated, some or all of the grid data points can be used as descriptors in generating QSAR models and analysing structure–activity relationships. Along with MFA descriptors, we used physicochemical and structural descriptors for obtaining the final QSAR models.

In the case of the RSA, the major steps are as follows: (1) generating conformers and energy minimisation; (2) aligning molecules using the MCSG method; (3) generating the receptor model based on the five most active compounds; (4) evaluating the compounds in the generated receptor model and (5) generating equations by the G/PLS method. RSA is a useful tool in situations when the 3D structure of the receptor is unknown [48] since one can build a hypothetical model of the receptor site. RSA differs from pharmacophore models in that the former tries to capture essential information about the receptor, while the latter captures information about the commonality of compounds that bind to a receptor. A receptor surface model (RSM) embodies essential information about the hypothetical receptor site as a 3D surface with associated properties such as hydrophobicity, partial charge, electrostatic potential, van der Waals (VDW) potential and hydrogen bonding propensity. The surface points that organise as triangular meshes in the construction of the RSA store these properties as associated scalar values. RSMs provide compact, quantitative descriptors, which capture 3D information of interaction energies in terms of steric and electrostatic fields at each surface point, which in other techniques such as CoMFA are calculated using probe interactions at various grid points. These descriptors can be used for 3D-QSAR studies. In the case of the RSA, we developed the QSAR equation taking physicochemical and structural descriptors along with RSA descriptors.

In the case of QSAR models with physicochemical descriptors and MSA models, the GFA technique [49,50] was used to generate a population of equations rather than one single equation for correlation between the biological activity and physicochemical properties. GFA involves the combination of multivariate adaptive regression splines algorithm with genetic algorithm to evolve a population of equations that best fit the training set data. It provides an error measure, called the lack of fit (LOF) score that automatically penalises models with too many features. It also inspires the use of splines as a powerful tool for nonlinear modelling. A distinctive feature of the GFA is that it produces a population of models (e.g. 100), instead of generating a single model, as do most other statistical methods. The range of variations in this population gives added information on the quality of fit and importance of the descriptors. It also inspires the use of splines as a powerful tool for nonlinear modelling. Splines are interpreted as performing either range identification or outlier removal.

In the case of the MFA and RSA, G/PLS was used as the statistical tool. The G/PLS algorithm [51,52] may be used as an alternative to a GFA calculation. G/PLS is derived from two QSAR calculation methods: GFA and PLS. The G/PLS algorithm uses GFA to select appropriate basis functions to be used in a model of the data and PLS regression as the fitting technique to weigh the basis functions' relative contributions in the final model. Application of G/PLS thus allows the construction of larger QSAR equations while still avoiding overfitting and eliminating most variables.

Statistical qualities

The statistical qualities of the equations were judged by the parameters such as determination coefficient (R^2) and variance ratio (F) at specified degrees of freedom (df) [53]. For G/PLS equations, least-squares error (LSE) was taken as an objective function to select an equation, while LOF

was noted for the GFA-derived equations. The generated QSAR equations were validated by leave-one-out (LOO) cross-validation R^2 (Q^2) and predicted residual sum of squares (PRESS) [54–56], and then were used for the prediction of enzyme inhibition activity values of the test set compounds. The prediction qualities of the models were judged by statistical parameters such as predictive R^2 (R^2_{pred}), squared correlation coefficient between observed and predicted values of the test set compounds with (r^2) and without (r^2_0) the intercept. It was previously shown that use of R^2_{pred} and r^2 might not be sufficient to indicate the external validation characteristics [57]. Thus, an additional parameter $r^2_{m(\text{test})}$ (defined as $r^2(1-\sqrt{r^2-r^2_0})$), which penalises a model for large differences between observed and predicted values of the test set compounds, was also calculated. Two other variants [58] of the r^2_m parameter, $r^2_{m(\text{LOO})}$ and $r^2_{m(\text{overall})}$, were also calculated. The parameter $r^2_{m(\text{overall})}$ is based on the prediction of both training (LOO prediction) and test set compounds. It was previously shown [58] that $r^2_{m(\text{LOO})}$ and $r^2_{m(\text{test})}$ penalise a model more strictly than Q^2 and R^2_{pred}, respectively. As an additional tool for validation, randomisation test was applied on the model development process and the developed models. The randomisation tests were performed at 90% confidence level for the process and the developed model were subjected to the randomisation test at 99% confidence level. Another parameter $R^2_p(R^2_p = R^2\sqrt{R^2-R^2_0})$ (R^2_{r} being squared mean correlation coefficient of random models) was also calculated [59,60] to check whether the models thus developed are not obtained by chance.

Results and discussion

The test set size was set to approximately 25% to the total data-set size [61], and the test set members are shown in Table 1. Statistical qualities of all important models are listed in Table 4. All the models have a Q^2 and R^2_{pred} value greater than 0.5 and the difference between R^2 and corresponding Q^2 values for the models is not very high (less than 0.3) [62].

Classical type QSAR (LFER model)

Equation (1) was among the best ones obtained from the GFA (5000 iterations). Both linear and spline terms were used to develop the equations,

$$\begin{aligned} p\text{IC}_{50} &= 2.541(\pm 0.102) - 0.507(\pm 0.101)\langle L_p - 2.74\rangle \\ &\quad + 0.562(\pm 0.124)I_{\text{Imidazole}} - 0.655(\pm 0.102)\pi_p \\ &\quad - 0.420(\pm 0.127)\pi_m \\ n_{\text{Training}} &= 20, \quad \text{LOF} = 0.132, \quad R^2 = 0.817, \\ R^2_{\text{a}} &= 0.770, \quad F = 16.70(\text{df}\,4, 15), \quad Q^2 = 0.699, \\ \text{PRESS} &= 1.313, \quad n_{\text{Test}} = 7, \quad R^2_{\text{pred}} = 0.621, \\ r^2_{m(\text{test})} &= 0.577, \quad r^2_{m(\text{LOO})} = 0.525, \quad r^2_{m(\text{overall})} = 0.530. \end{aligned} \tag{1}$$

The standard errors of regression coefficients are given within parentheses. Equation (1) could explain 77.0% of the variance while it could predict 69.9% of the variance (LOO-predicted variance). When the equation was used to predict the CYP17 inhibition potency of the test set compounds, the predicted R^2 (R^2_{pred}) value was found to be 0.621. The r^2_m values for the test, training and overall sets were statistically significant ($r^2_{m(\text{test})} = 0.577$, $r^2_{m(\text{LOO})} = 0.525$ and $r^2_{m(\text{overall})} = 0.530$). The relative importance of the descriptors according to their standardised regression coefficient values is of the following order: $\langle L_p - 2.74\rangle > I_{\text{Imidazole}} > L_p - 2.74\rangle > I_{Imidazole} > \pi_p > \pi_m$.

L is a steric parameter, which is the length of the substituents along the axis of its substitution to the parent skeleton. The term $\langle L_p - 2.74\rangle$ with negative regression coefficient indicates that to avoid detrimental interactions, the value of L_p should be less than 2.74. L_p is the length of the substitutions at the 4 (*para*) position of the terminal phenyl ring. It has been observed that H or —OH substitution at 4 position of the terminal phenyl ring (such as **1**, **3** and **21**) showed more activity due to their small L value than other substitutions (such as —SMe in the case of compound **8**).

$I_{\text{Imidazole}}$ is an indicator parameter denoting the pattern of attachment of the imidazole group to the biphenyl ring. It is 1 and 0 for the *para* and *meta* positions, respectively, of the imidazole group with respect to the second phenyl group. The positive coefficient of this parameter indicates that the *para* substitution of the imidazole group is favourable for the activity.

Table 4. Comparison of statistical qualities of different models.

Type of models	R^2	Q^2	R^2_{pred}	$r^2_{m(\text{test})}$	$r^2_{m(\text{LOO})}$	$r^2_{m(\text{overall})}$
LFER	0.817	0.699	0.621	0.577	0.525	0.530
MSA	0.915	0.827	0.808	0.788	0.674	0.708
MFA	0.910	0.809	0.876	0.792	0.785	0.820
RSA	0.954	0.932	0.853	0.763	0.908	0.858

The parameter π is the lipophilicity substitution constant, which is a very important parameter in modelling studies. Both the terms π_m and π_p have negative regression coefficient. The negative coefficient of the lipophilic substituent constants (π_m and π_p) of the *meta* and *para* substituents of the terminal phenyl ring indicates that lipophilic substitutions are detrimental for the inhibitory activity. Compounds with a hydrophilic (—OH) substituent at the *para* position of the terminal phenyl ring (such as compounds **3** and **21**) showed good inhibitory potency when compared to compounds with —Cl, —Me, —SMe and —F substitutions such as compounds **5**, **6**, **8** and **22**, respectively. A similar trend is also observed in the case of the *meta* position of the terminal phenyl ring. For example, —OH and —NH_2 substituents (compounds **10** and **16**) are more favourable for the activity than —Cl and —F (in the case of compounds such as **12** and **11**, respectively).

Molecular shape analysis

A view of the aligned training set molecules is shown in Figure 2. The best equation is obtained from the GFA (5000 iterations) with linear terms,

$$\begin{aligned} p\mathrm{IC}_{50} = {} & 12.795(\pm 0.994) - 0.607(\pm 0.062)\mathrm{Shadow\_Xlength} \\ & - 0.490(\pm 0.098)\mathrm{Rotatablebonds} \\ & - 0.075(\pm 0.013)\mathrm{Shadow\_YZ} \\ & + 0.339(\pm 0.069)\mathrm{Hbonddonor} \end{aligned} \quad (2)$$

$n_{\mathrm{Training}} = 21,\ \mathrm{LOF} = 0.050,\ R^2 = 0.915,$

$R^2_{\mathrm{a}} = 0.894,\ F = 43.22(\mathrm{df}\,4,16),\ Q^2 = 0.827,$

$\mathrm{PRESS} = 0.825,\ n_{\mathrm{Test}} = 7,\ R^2_{\mathrm{pred}} = 0.808,$

$r^2_{m(\mathrm{test})} = 0.788,\ r^2_{m(\mathrm{LOO})} = 0.674,\ r^2_{m(\mathrm{overall})} = 0.708.$

According to the standardised regression coefficients, the relative importance of the descriptors in Equation (2) is in the following order: Shadow_Xlength > Rotatablebonds > Shadow_YZ > Hbonddonor. The standard errors of regression coefficients are given within parentheses. The equation was found to be statistically significant with an explained variance of 89.4% and a LOO-predicted variance of 82.7%. When the equation is applied on the test set compounds, the R^2_{pred} value was found to be 0.808. Statistical significance of the model was also indicated by r^2_m parameters which are listed in Table 4.

Shadow_Xlength (Lx) is the length of the molecule in the *X* dimension. The negative regression coefficient of the term indicates that an increase in the length of the molecule in the *X* dimension (Lx) [for example, in the case of compounds **2**, **7**, **8** and **26** (imidazole-substituted biphenyl derivatives with *p*-OCH_3, *p*-CN, *p*-SMe, *m*-$NHCOCH_3$ groups at the terminal phenyl ring)] is detrimental for the inhibitory activity. Similarly, compounds with *p*-OH, *m*-OH, *p*-NH_2 groups at the terminal phenyl ring (compounds **3**, **10** and **16**) or replacement of the terminal phenyl ring by the thiophene ring (in the case

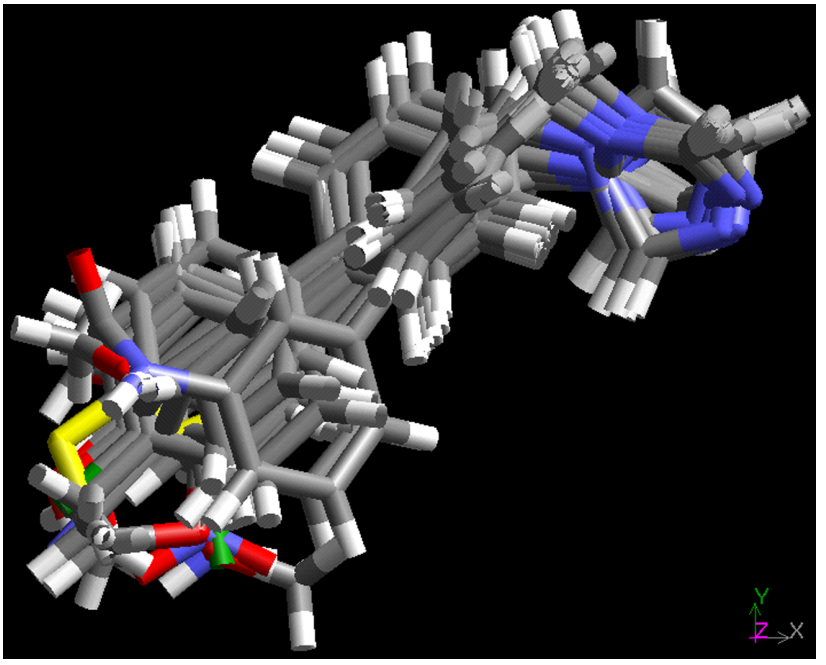

Figure 2. Aligned geometry of the training set molecules.

of compound **28**) showed higher activity due to lower values of Lx.

Rotatablebonds is the number of rotatable bonds in the molecules, and the negative coefficient of the term indicates importance of restriction of molecular flexibility. Compounds **16** and **28** with less number of rotatable bonds and shadow parameters (Lx, Syz) value in the lower range showed good inhibitory activity. Compounds **24**–**26** with more number of rotatable bonds showed poor inhibitory activity. Compounds with less number of rotatable bonds such as **1**, **5**, **6** and **12** showed poor inhibitory activity due to relatively large value of the two shadow parameters (Shadow_Xlength, Shadow_YZ) and without any hydrogen bond donor group.

Shadow_YZ (Syz) is defined as the area of the molecular shadow in the *YZ* plane. The negative coefficient of the term indicates that an increase in the area of the molecular shadow in the *YZ* plane (such as compounds **24** and **26**) is detrimental for the activity. Compound **28** with least value of the parameter showed good inhibitory activity. It has been observed that compounds with moderate values of the parameters (Lx, Syz and Rotatablebonds) with at least one hydrogen bond donor group (such as compounds **3**, **10** and **16**) show good inhibitory activity.

The positive coefficient of Hbonddonor indicates that compounds with more number of hydrogen bond donor groups showed good inhibitory activity (**3**, **10** and **16**). Compounds **25** and **26** showed poor inhibitory activity in spite of having one hydrogen bond donor group due to high values of other parameters (Lx, Syz and Rotatablebonds).

Molecular field analysis

The field generated was of the rectangular type. The probes used were H^+, CH_3, CH_3^-, CH_3^+ and donor/acceptor. The charge method used was the Gasteiger type, and the energy cut-off was kept at -50 to $+50$ kcal. Additionally, we added physicochemical and structural descriptors as the independent variables along with the MFA descriptors. The QSAR equation was generated using the G/PLS method (1000 crossovers, linear terms, scaled variables, number of components 2, initial equation length 2, no fixed length of the final equation and other default settings). A view of the aligned training set molecules in the field is shown in Figure 2. The following equation was obtained from

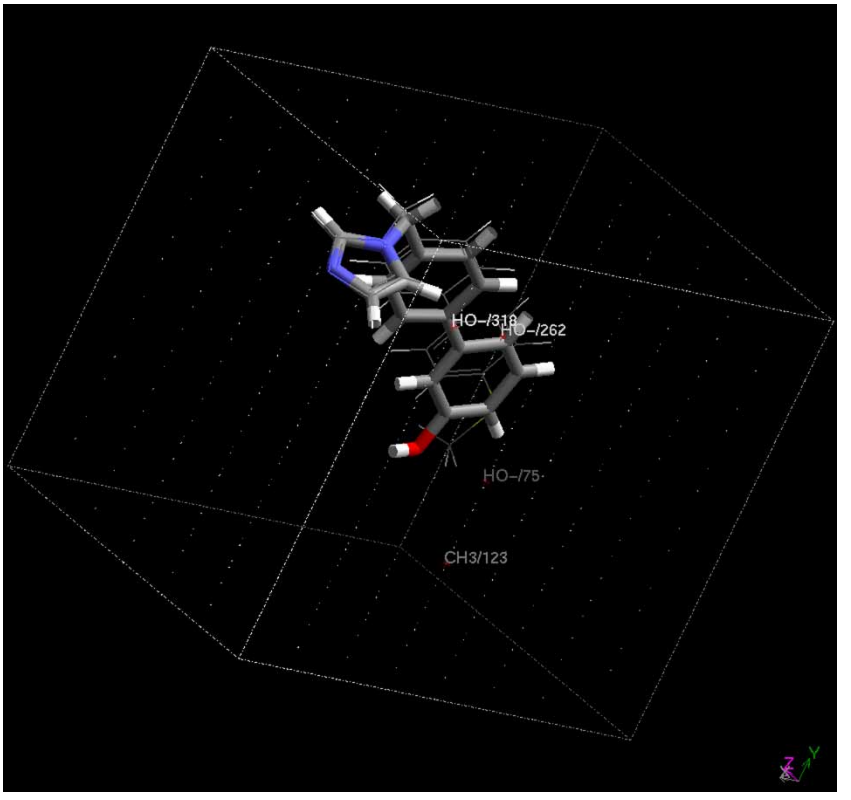

Figure 3. The view of the most active compound **10** (cylinder model) and least active compound **8** (stick model) in the molecular field with important interaction points.

the MFA:

$$p\mathrm{IC}_{50} = 4.673 - 0.010\langle \mathrm{HO}^-/75 + 0.716\rangle - 0.010\langle \mathrm{HO}^-/262 + 0.670\rangle - 0.409\langle 1 - \mathrm{Hbonddonor}\rangle - 0.009\mathrm{CH}_3/123 - 0.008\langle \mathrm{HO}^-/318 + 0.021\rangle - 0.318A\log P \quad (3)$$

$n_{\mathrm{Training}} = 21$, LSE $= 0.020$, $R^2 = 0.910$, $R^2_{\mathrm{a}} = 0.900$, $F = 90.79(\mathrm{df}\,2,18)$, $Q^2 = 0.809$, PRESS $= 0.908$, $n_{\mathrm{Test}} = 7$, $R^2_{\mathrm{pred}} = 0.876$, $r^2_{m(\mathrm{test})} = 0.792$, $r^2_{m(\mathrm{LOO})} = 0.785$, $r^2_{m(\mathrm{overall})} = 0.820$.

The relative order of importance of the descriptors is as follows: $\langle \mathrm{OH}^-/75 + 0.716\rangle > \langle \mathrm{OH}^-/262 + 0.670\rangle > \langle 1 - \mathrm{Hbonddonor}\rangle > \mathrm{CH}_3/123 > \langle \mathrm{OH}^-/318 + 0.021\rangle > A\log P$.

Equation (3) could explain and predict, respectively, 90.0 and 80.9% of the variance. The external validation statistics of Equation (3) are very good ($R^2_{\mathrm{pred}} = 0.876$ and $r^2_{m(\mathrm{test})} = 0.792$). In Equation (3), $\mathrm{OH}^-/318$, $\mathrm{OH}^-/262$, $\mathrm{CH}_3/123$ and $\mathrm{OH}^-/75$ are the probes which correspond to the spatial position as shown in Figure 3. These terms indicate the interactions of the probe with the ligands at points 318, 262, 123 and 75, respectively. Among the above points, 75 and 123 are close to the *para* position of the terminal phenyl ring and points 318 and 262 are near the imidazole ring.

The negative coefficient of the term $\langle \mathrm{OH}^-/75 + 0.716\rangle$ indicates that substitutions at the terminal phenyl ring reduce the interaction energy between the hydroxy probe at the grid position and the ligand and facilitate the inhibitory activity. Lipophilic and or bulky substitutions at this point are detrimental for the activity (such as compounds **2**, **6** and **8**). Hydrophilic substitutions with hydrogen bonding capacity lower the interaction energy between the probe and the target molecule and increase the inhibitory activity (such as compounds **3**, **10** and **16**). The parameter $\mathrm{CH}_3/123$ with negative regression coefficient indicates that the presence of lipophilic and/or bulky groups at the terminal phenyl ring is detrimental for the activity (compound **8**).

It has been observed that the imidazole group present at the *meta* position of the terminal phenyl ring interacts with points 318 and 262. However, for the imidazole group present at the *para* position of the terminal phenyl ring, points are far away from the imidazole group. The negative coefficients of the terms $\langle \mathrm{OH}^-/262 + 0.670\rangle$ and $\langle \mathrm{OH}^-/318 + 0.021\rangle$ thus indicate the optimum position of the imidazole moiety (i.e. *para*) for favourable inhibitory activity.

In Equation (2), Hbonddonor has positive coefficient and in Equation (3), $\langle 1 - \mathrm{Hbonddonor}\rangle$ has negative regression coefficient. This indicates that for the optimal inhibitory activity, the number of hydrogen bond donors should be more than 1 or at least 1 (compounds **3**, **10** and **16**).

The negative coefficient of $A\log P$ (log of partition coefficient) indicates that compounds with high values of the parameter (such as compounds **5**, **6**, **12** and **22**) show poor inhibitory activity than compounds with less value of the parameter (compounds **10**, **16** and **28**). Compound **26** having the least value of the parameter shows poor inhibitory activity due to the presence of the imidazole ring at the *meta* position of the terminal phenyl ring as stated earlier.

Receptor surface analysis

The RSM is usually generated from the most active compounds in the data-set. The rationale is that the most active molecules tend to explore the best spatial and electronic interactions with the receptor, while the least active molecules do not and tend to have unfavourable steric and electrostatic interactions. We initiated our study with receptor models generation using the five best active compounds (compounds **3**, **9**, **10**, **16** and **28**) in the training set. After the generation of the RSM, all the structures in the training and test sets can be evaluated against the model. The model can be used to calculate the energy associated with the binding of a molecule in the model. It can also be used to minimise a molecule by adjusting the geometry of the structure into a 'best-fit configuration' based on the constraints imposed by the receptor model. The best model (G/PLS, 1000 crossovers, linear and spline terms, number of components 2, initial equation length 2 and other default settings) obtained from the training set compounds is as follows:

$$p\mathrm{IC}_{50} = 3.002 + 35.351\langle -0.617 - \mathrm{TOT}/494\rangle - 0.408\langle 1 - \mathrm{Hbonddonor}\rangle - 0.582\langle A\log P98 - 3.025\rangle - 0.852\langle -0.085 - \mathrm{TOT}/2094\rangle \quad (4)$$

$n_{\mathrm{Training}} = 21$, LSE $= 0.010$, $R^2 = 0.954$, $R^2_{\mathrm{a}} = 0.949$, $F = 185.66(\mathrm{df}\,2,18)$, $Q^2 = 0.932$, PRESS $= 0.323$, $n_{\mathrm{Test}} = 7$, $R^2_{\mathrm{pred}} = 0.853$, $r^2_{m(\mathrm{test})} = 0.763$, $r^2_{m(\mathrm{LOO})} = 0.908$, $r^2_{m(\mathrm{overall})} = 0.858$.

The relative importance of the descriptors according to their standardised value is of the

following order: $\langle -0.617 - \mathrm{TOT}/494\rangle > \langle 1 - \mathrm{Hbonddonor}\rangle > \langle A\log P98 - 3.025\rangle > \langle -0.085 - \mathrm{TOT}/2094\rangle$. Equation (4) could explain and predict, respectively, 94.9 and 93.2% of the variance. The external validation statistics of Equation (4) are very good. The predictive R^2 value for the test set was 0.853 and r_m^2 values for the test, training and overall sets were found to be 0.763, 908 and 0.858, respectively.

The descriptors TOT/494, TOT/2094 are the sum of both electrostatic interaction energy and VDW interaction energy at points 494 and 2094, respectively. By using receptor data to develop a QSAR model, the goodness of fit can be evaluated between a candidate structure and a postulated pseudoreceptor. The view of the RSM embedded with most active compound **10** and compound **25** and the interaction points is shown in Figures 4–6, mapped with charge, hydrogen bonding and hydrophobicity, respectively. The RSM represents essential information about the hypothetical receptor site as a 3D surface with associated properties mapped onto the surface model. The location and magnitude of a descriptor can be used as a guideline to improve the activity of molecules. The surface represents information about the steric nature of the receptor site and the associated properties of interest, such as hydrophobicity, partial charge and hydrogen-bonding propensity. In Figure 4, the red areas are positively charged, blue areas are negatively charged and white areas are neutral. Similarly, in Figure 5, purple areas act as hydrogen bond donors, light blue areas act as hydrogen bond acceptors and white areas do not have hydrogen bonding activity. The brown areas in Figure 6 are hydrophobic. Figures 7 and 8 are the RSM embedded with the most active compound **10** and the least active compound **8** mapped with total energy, respectively. The magenta colour represents favourable interaction sites, while the green-coloured regions represent unfavourable sites for the binding of the molecule on the receptor surface. From Figure 7, it can be observed that favourable interactions occur in the case of compound **10** near the phenyl ring with —OH substitution, while Figure 8 shows that substitution such as —SMe at the phenyl ring (compound **8**) causes detrimental interactions. Favourable interactions can be seen near the imidazole ring. Figures 5 and 6 showing RSMs mapped with hydrogen bond donor property and hydrophobicity, respectively, suggest that the surface near the substitution on the terminal phenyl ring is hydrophilic and hydrogen bond donor groups are present at this position. Therefore, it can be inferred that hydrophilic and hydrogen bond acceptor groups are favoured as substitutions in the terminal phenyl ring for the inhibitory activity. The surface near the imidazole ring is moderately hydrophilic and hydrogen bond acceptor groups are present in this site.

Point 494 is near the imidazole ring and interaction at this point is not favourable for receptor–ligand interactions. The negative coefficient of the term $\langle -0.617 - \mathrm{TOT}/494\rangle$ indicates that the interaction value at point 494 should be less than -0.617 for the inhibitory activity. Compounds such as **9**, **10**, **16** and **28** (*para*-substituted compounds) with low TOT/494 value (or high negative value) showed good

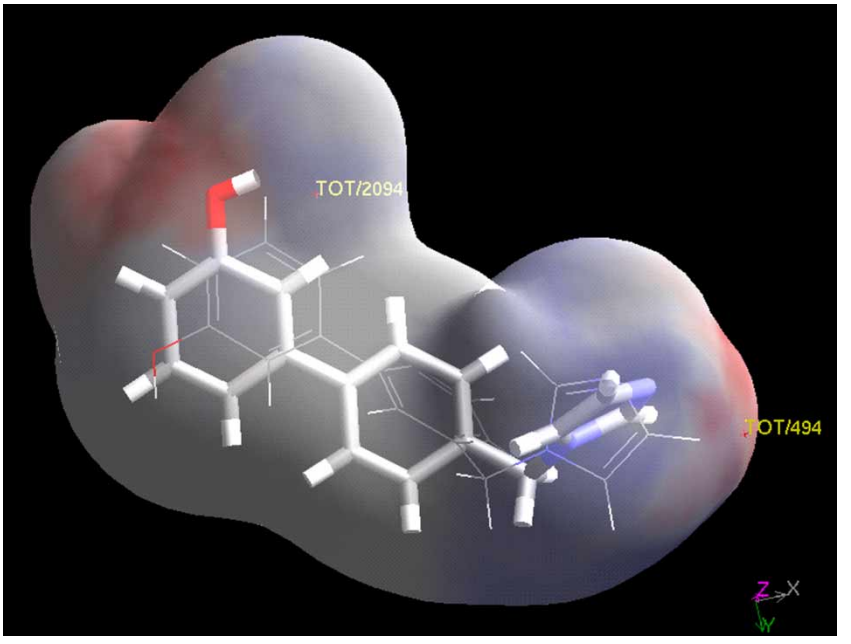

Figure 4. The most active compound **10** (cylinder model) and compound **25** (stick model) embedded into the RSM mapped with charge: the red areas are positively charged, blue areas are negatively charged and white areas are neutral.

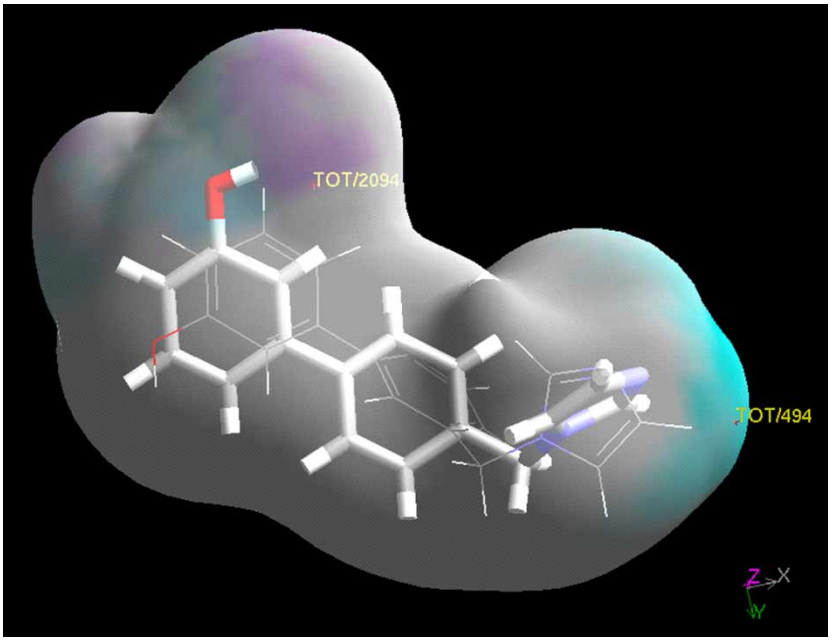

Figure 5. The most active compound **10** (cylinder model) and compound **25** (stick model) embedded into the RSM mapped with hydrogen bonding: the purple areas act as hydrogen bond donors, light blue areas act as hydrogen bond acceptors and white areas do not have hydrogen bonding activity.

inhibitory activity. On the other hand, compounds **19**, **22** and **25** (*meta*-substituted biphenyls) showed poor activity due to high TOT/494 value.

The number of hydrogen bond donor groups for inhibitory activity should be 1 or more than 1 as stated in Equation (3).

Point 2094 is close to the terminal phenyl ring and interaction at this point is favourable for receptor binding for good inhibitory activity. Compound **19** with high interaction value at point 2094 showed poor inhibitory activity due to the lack of a hydrogen bond donor group as well as high interaction energy value at point 494.

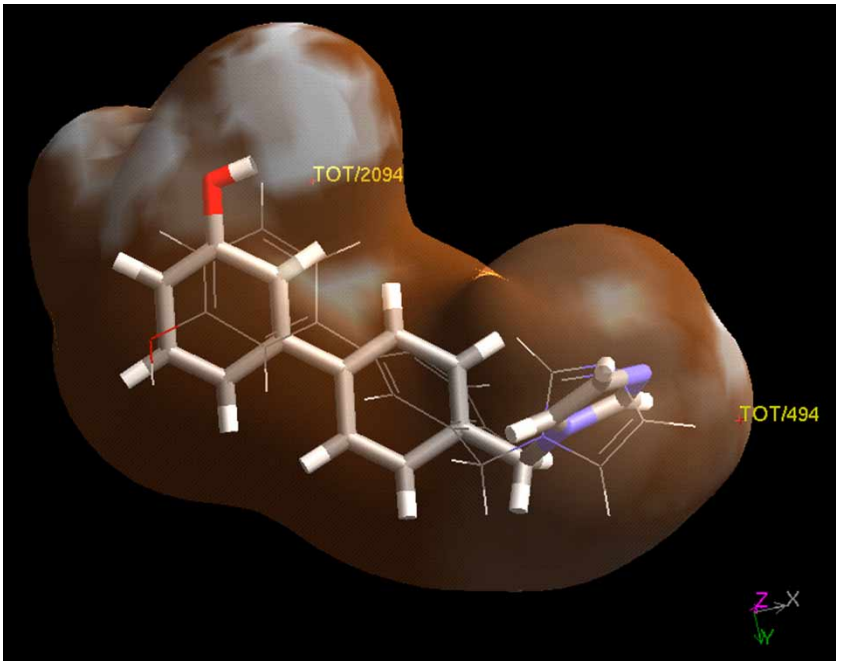

Figure 6. The most active compound **10** (cylinder model) and compound **25** (stick model) embedded into the RSM mapped with hydrophobicity: the brown areas are hydrophobic.

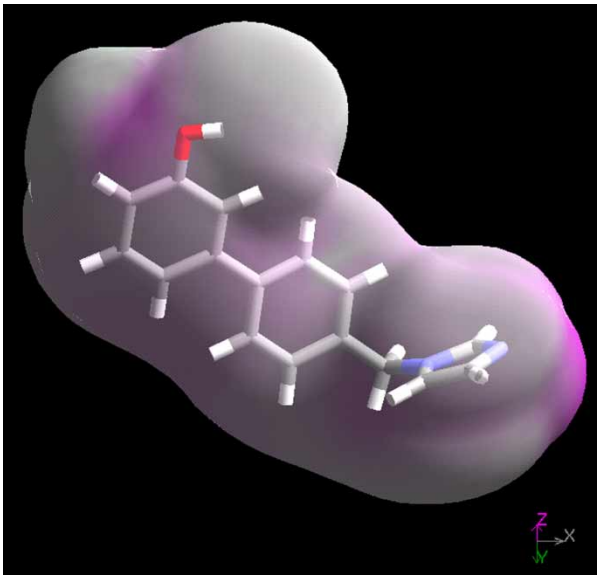

Figure 7. The best active compound **10** embedded into the RSM mapped with total energy: the magenta colour represents favourable interaction sites.

Compounds with interaction energy at higher range at point 2094 (such as **3**, **10** and **16**) with at least 1 or more hydrogen bond donor groups showed better inhibitory activity than compounds (such as **24**–**26**) with lower interaction energy at point 2094.

The value of $A\log P98$ should be less than 3.025 for optimum inhibitory activity as the term $\langle A\log P98 - 3.025\rangle$ bears negative regression coefficient. Compounds with lower value of $A\log P98$ (such as **3**, **10**, **16** and **28**) showed good inhibitory activity than highly lipophilic compounds (such as **5**, **6**, **8** and **12**).

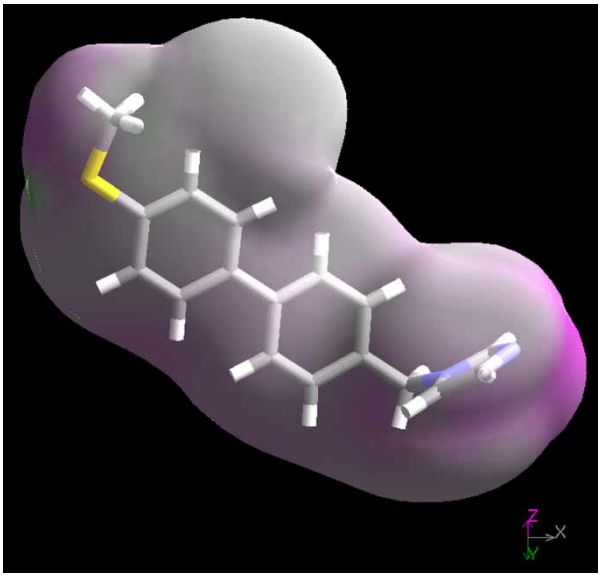

Figure 8. The least active compound **8** embedded into the RSM mapped with total energy: the magenta colour represents favourable interaction sites, while the green colour region represents unfavourable sites for binding of the molecule to the receptor surface.

Randomisation

Further validation of the models was carried out using the *Y* scrambling technique. The process randomisation test has been performed at 90% confidence level and the developed models were subjected to the randomisation test at 99% confidence interval. The *Y* column was permuted randomly and the squared average correlation coefficient (R_r^2) of all the randomised models was calculated. The process randomisation is different from the model randomisation in that the descriptor selection process is repeated from the whole pool of descriptors in the former case, while, in the latter case, only those descriptors present in the model are used. The values of R_r^2 obtained for all the models were significantly lower than the squared correlation coefficient (R^2) of the non-randomised model except for the MFA and RSA process randomisation. Since no guideline is given as to how much this difference should be, the values of R_p^2 were calculated for all the developed models. This parameter penalises the model R^2 for small difference between R^2 and R_r^2. Table 5 gives a detailed report of the different results obtained in the randomisation test. The values of R_p^2 obtained for all the models (model randomisation) were quite above the acceptable value of 0.5 (0.638–0.926). The results thus infer that all the models developed were robust enough and not the outcome of mere chance. The process randomisation results of MFA and RSA did not cross the required value of 0.5.

Comparison of the present models with previously published QSARs on the present data-set

In Table 6, statistical quality of the models developed in this paper is compared to those reported previously [40]. However, in the previous paper [40], less number of compounds was considered for model development; thus, a direct comparison is not possible.

Table 5. Randomisation results for the process and models.

Type of models	Type of randomisation	R^2	R_r^2	R_p^2
LFER	Process (90%)	0.817	0.426	0.511
	Model (99%)	0.817	0.207	0.638
MSA	Process (90%)	0.915	0.368	0.676
	Model (99%)	0.915	0.193	0.778
MFA	Process (90%)	0.910	0.640	0.473
	Model (99%)	0.910	0.005	0.866
RSA	Process (90%)	0.954	0.755	0.425
	Model (99%)	0.954	0.012	0.926

Table 6. Comparison of the present models with previous QSARs on this data-set.

Reference	Modelling techniques	No. of descriptors	No. of compounds	R^2
Wachall et al. [40]	Linear	2	8	0.941
	Linear	2	8	0.884
	Linear	3	16	0.903
Present study	GFA, Equation (1)	4	20/7[a]	0.817
	GFA, Equation (2)	4	20/7[a]	0.915
	G/PLS, Equation (3)	6	20/7[a]	0.910
	G/PLS, Equation (4)	4	20/7[a]	0.954

[a] The number of training set compounds is 20 and the number of test set compounds is 7.

Overview and conclusions

The whole data-set of 28 compounds was divided into a training set and a test set based on *K*-means clustering of the standardised physicochemical descriptor matrix and models were developed from the training set. The predictive ability of the models was judged from the prediction of the CYP17 inhibition activity of the test set compounds. A comparison of statistical quality of different models is given in Table 4. All the models suggest that the presence of the imidazole group at the *para* position of the biphenyl ring is favourable for the inhibitory activity which is in agreement with the source paper [40]. All the models (except MSA) indicate negative contribution of lipophilicity of compounds towards inhibitory activity. QSAR analysis with physicochemical descriptors indicates the importance of steric descriptors (L_p) related to the length of substituents at the *para* position of the terminal phenyl ring as well as lipophilic substituents (π_m and π_p) constant at the terminal phenyl ring. The length of the substitution at the *para* position of the phenyl ring should be small (—H, —OH) as in the case of compounds **1**, **3** and **21** to avoid steric interactions as compared to other substitution such as —SMe (compound **8**). Substitutions such as —OH and —NH_2 at the terminal phenyl ring will retain inhibitory activity and —Cl, —Me, —SMe and —F substitutions will disfavour inhibitory activity. MSA analysis indicates the importance of shadow indices (Shadow_Xlength, Shadow_YZ) and hydrogen bond donor groups as well as molecular flexibility (Rotatablebonds). The results indicate that small substituents (—OH, —NH_2) at the terminal phenyl ring are favoured when compared to other substitutions (—OCH_3, —CN, —SMe, —$NHCOCH_3$). MFA results showed specific interactions of the probes at different grid locations with the molecules (thus indicating importance of different substituents on the terminal phenyl ring and the imidazole ring). Hydrophilic substitutions with hydrogen bonding capacity at the terminal phenyl ring (*meta* or *para* position) are conducive for inhibitory activity than bulky or lipophilic substitutions. The RSM gives information about the steric nature of the receptor site and the associated properties of interest, such as hydrophobicity, hydrophilicity and hydrogen bonding propensity. Overall, it can be concluded that the substitutions at the terminal phenyl ring should be small and hydrophilic with hydrogen bonding capacity. There is no crystal structure available for the human CYP17 enzyme (homology model is available) [23,30,31]. In computational modelling, ligand-based approaches are the main alternative for those enzymes without crystal structures. Thus, graphical representations of the beneficial and non-beneficial interactions in the MFA will allow a medicinal chemist to design new structures. Knowledge of receptor properties of interest can be extracted from the graphs derived in 3D-QSAR. Shape and physicochemical properties of substitutions along with the above knowledge can also be identified for the favourable receptor–drug interaction. Randomisation results for the process and models are listed in Table 5. Based on the model randomisation, the R_p^2 values for the developed models are well above the cut-off value (>0.5). It has been observed that G/PLS-derived models are superior to GFA-derived models based on validation parameters. The RSA-derived G/PLS model was found to be the best model based on the highest internal ($Q^2 = 0.932$) predictive power as well as $r^2_{m(\text{overall})}$ criteria (0.858). According to external prediction statistics ($R^2_{\text{pred}} = 0.876$), the MFA-derived G/PLS model outperforms the other models. However, based on the process randomisation, the MSA model is more reliable than the other equations.

Acknowledgement

One of the authors (PPR) thanks the University Grants Commission (UGC), New Delhi for a fellowship.

References

[1] D.M. Parkin, *Global cancer statistics in the year 2000*, Lancet Oncol. 2 (2001), pp. 533–543.
[2] A. Jemal, R. Siegel, E. Ward, T. Murray, J. Xu, and M.J. Thun, *Cancer statistics, 2007*, CA Cancer J. Clin. 57 (2007), pp. 43–66.
[3] A. Jemal, T. Murray, E. Ward, A. Samuels, R.C. Tiwari, A. Ghafoor, E.J. Feuer, and M.J. Thun, *Cancer statistics, 2005*, CA Cancer J. Clin. 55 (2005), pp. 10–30.
[4] C. Huggins and C.V. Hodges, *Studies on prostatic cancer. I. The effect of castration, of estrogen and of androgen injection on serum phosphatases in metastatic carcinoma of the prostate*, Cancer Res. 1 (1941), pp. 293–297.
[5] C. Huggins, R.E. Stevens, Jr, and C.V. Hodges, *Studies on prostatic cancer: Ii. The effects of castration on advanced carcinoma of the prostate gland*, Arch. Surg. 43 (1941), pp. 209–223.
[6] I. Huhtaniemi, H. Nikula, M. Parvinen, and S. Rannikko, *Histological and functional changes of the testis tissue during GnRH agonist treatment of prostatic cancer*, Am. J. Clin. Oncol. 11 (Suppl 1) (1988), pp. S11–S15.
[7] J.A. Smith, *New methods of endocrine management of prostatic cancer*, J. Urol. 137 (1987), pp. 1–10.

[8] E.D. Crawford, M.A. Eisenberger, D.G. McLeod, J.T. Spaulding, R. Benson, F.A. Dorr, B.A. Blumenstein, M.A. Davis, and P.J. Goodman, *A controlled trial of leuprolide with and without flutamide in prostatic carcinoma*, N. Engl. J. Med. 321 (1989), pp. 419–424.

[9] C.D. Chen, D.S. Welsbie, C. Tran, S.H. Baek, R. Chen, R. Vessella, M.G. Rosenfeld, and C.L. Sawyers, *Molecular determinants of resistance to antiandrogen therapy*, Nat. Med. 10 (2004), pp. 33–39.

[10] B.A. Hellerstedt and K.J. Pienta, *The current state of hormonal therapy for prostate cancer*, CA Cancer J. Clin. 52 (2002), pp. 154–179.

[11] R. Arimoto, *Computational models for predicting interactions with cytochrome p450 enzyme*, Curr. Top. Med. Chem. 6 (2006), pp. 1609–1618.

[12] E.M. Sellers, H.L. Kaplan, and R.F. Tyndale, *Inhibition of cytochrome P450 2A6 increases nicotine's oral bioavailability and decreases smoking*, Clin. Pharmacol. Ther. 68 (2000), pp. 35–43.

[13] F.J. Gonzalez, *Molecular genetics of the P-450 superfamily*, Pharmacol. Ther. 45 (1990), pp. 1–38.

[14] F.P. Guengerich, *Characterization of human cytochrome P450 enzymes*, FASEB J. 6 (1992), pp. 745–748.

[15] S. Rendic, *Summary of information on human CYP enzymes: Human P450 metabolism data*, Drug Metabol. Rev. 34 (2002), pp. 83–448.

[16] R.D. Bruno and V.C. Njar, *Targeting cytochrome P450 enzymes: A new approach in anti-cancer drug development*, Bioorg. Med. Chem. 15 (2007), pp. 5047–5060.

[17] B.C. Chung, J. Picado-Leonard, M. Haniu, M. Bienkowski, P.F. Hall, J.E. Shively, and W.L. Miller, *Cytochrome P450c17 (steroid 17 alpha-hydroxylase/17,20 lyase): Cloning of human adrenal and testis cDNAs indicates the same gene is expressed in both tissues*, Proc. Natl Acad. Sci. USA 84 (1987), pp. 407–411.

[18] M.K. Akhtar, S.L. Kelly, and M.A. Kaderbhai, *Cytochrome b(5) modulation of 17{alpha} hydroxylase and 17-20 lyase (CYP17) activities in steroidogenesis*, J. Endocrinol. 187 (2005), pp. 267–274.

[19] N.W. Kolar, A.C. Swart, J.I. Mason, and P. Swart, *Functional expression and characterisation of human cytochrome P45017alpha in* Pichia pastoris, J. Biotechnol. 129 (2007), pp. 635–644.

[20] K.A. Harris, V. Weinberg, R.A. Bok, M. Kakefuda, and E.J. Small, *Low dose ketoconazole with replacement doses of hydrocortisone in patients with progressive androgen independent prostate cancer*, J. Urol. 168 (2002), pp. 542–545.

[21] J. Eklund, M. Kozloff, J. Vlamakis, A. Starr, M. Mariott, L. Gallot, B. Jovanovic, L. Schilder, E. Robin, M. Pins, and R.C. Bergan, *Phase II study of mitoxantrone and ketoconazole for hormone-refractory prostate cancer*, Cancer 106 (2006), pp. 2459–2465.

[22] R.A. Madan and P.M. Arlen, *Abiraterone. Cougar biotechnology*, IDrugs 9(1) (2006), pp. 49–55.

[23] U.E. Hille, Q. Hu, C. Vock, M. Negri, M. Bartels, U. Müller-Vieira, T. Lauterbach, and R.W. Hartmann, *Novel CYP17 inhibitors: Synthesis, biological evaluation, structure–activity relationships and modelling of methoxy- and hydroxy-substituted methyleneimidazolyl biphenyls*, Eur. J. Med. Chem. 44 (2009), pp. 2765–2775.

[24] J.G. Gambertoglio, W.J. Amend, Jr, and L.Z. Benet, *Pharmacokinetics and bioavailability of prednisone and prednisolone in healthy volunteers and patients: A review*, J. Pharmacokinet. Biopharm. 8 (1980), pp. 1–52.

[25] K. Fotherby, *Bioavailability of orally administered sex steroids used in oral contraception and hormone replacement therapy*, Contraception 54 (1996), pp. 59–69.

[26] J.F. Steiner, *Clinical pharmacokinetics and pharmacodynamics of finasteride*, Clin. Pharmacokinet. 30 (1996), pp. 16–27.

[27] A. Hameed, T. Brothwood, and P. Bouloux, *Delivery of testosterone replacement therapy*, Curr. Opin. Investig. Drugs 4 (2003), pp. 1213–1219.

[28] J.E. Buster, P.R. Casson, A.B. Straughn, D. Dale, E.S. Umstot, N. Chiamori, and G.E. Abraham, *Postmenopausal steroid replacement with micronized dehydroepiandrosterone: Preliminary oral bioavailability and dose proportionality studies*, Am. J. Obstet. Gynecol. 166 (1992), pp. 1163–1168.

[29] R.E. Peterson, *Metabolism of adrenal cortical steroids*, in *The Human Adrenal Cortex*, N.P. Christy, ed., Harper & Row, New York, 1997, pp. 87–189.

[30] Q. Hu, M. Negri, K. Jahn-Hoffmann, Y. Zhuang, S. Olgen, M. Bartels, U. Müller-Vieira, T. Lauterbach, and R.W. Hartmann, *Synthesis, biological evaluation, and molecular modeling studies of methylene imidazole substituted biaryls as inhibitors of human 17alpha-hydroxylase-17,20-lyase (CYP17) – part II: Core rigidification and influence of substituents at the methylene bridge*, Bioorg. Med. Chem. 16 (2008), pp. 7715–7727.

[31] C. Jagusch, M. Negri, U.E. Hille, Q. Hu, M. Bartels, K. Jahn-Hoffmann, M.A. Pinto-Bazurco Mendieta, B. Rodenwaldt, U. Müller-Vieira, D. Schmidt, T. Lauterbach, M. Recanatini, A. Cavalli, and R.W. Hartmann, *Synthesis, biological evaluation and molecular modelling studies of methyleneimidazole substituted biaryls as inhibitors of human 17alpha-hydroxylase-17,20-lyase (CYP17). Part I: Heterocyclic modifications of the core structure*, Bioorg. Med. Chem. 16 (2008), pp. 1992–2010.

[32] G.A. Wächter, R.W. Hartmann, T. Sergejew, G.L. Grün, and D. Ledergerber, *Tetrahydronaphthalenes: Influence of heterocyclic substituents on inhibition of steroid enzymes P450 arom and P450 17*, J. Med. Chem. 39 (1996), pp. 834–841.

[33] F.C. Chan, G.A. Potter, S.E. Barrie, B.P. Haynes, M.G. Rowlands, J. Houghton, and M. Jarman, *3- and 4-pyridylalkyl adamantanecarboxylates: Inhibitors of human cytochrome P450(17 alpha) (17 alpha-hydroxylase/C17,20-lyase). Potential nonsteroidal agents for the treatment of prostatic cancer*, J. Med. Chem. 39 (1996), pp. 3319–3323.

[34] R. McCague, M.G. Rowlands, S.E. Barrie, and J. Houghton, *Inhibition of enzymes of estrogen and androgen biosynthesis by esters of 4-pyridylacetic acid*, J. Med. Chem. 33 (1990), pp. 3050–3055.

[35] M.G. Rowlands, S.E. Barrie, F. Chan, J. Houghton, M. Jarman, R. McCague, and G.A. Potter, *Esters of 3-pyridylacetic acid that combine potent inhibition of 17 alpha-hydroxylase/C17,20-lyase (cytochrome P45017 alpha) with resistance to esterase hydrolysis*, J. Med. Chem. 38 (1995), pp. 4191–4197.

[36] R.W. Hartmann, G.A. Wächter, T. Sergejew, R. Würtz, and J. Düerkop, *4,5-Dihydro-3-(2-pyrazinyl)naphtho[1,2-c]pyrazole: A potent and selective inhibitor of steroid-17 alpha-hydroxylase-C17,20-lyase (P450 17)*, Arch. Pharm. (Weinheim) 328 (1995), pp. 573–575.

[37] A.M. Al-Hamrouni, M. Ahmadi, P.J. Nicholls, H.J. Smith, P. Lombardi, and V. Pestellini, *1-[(Benzofuran-2-yl)phenylmethyl] imidazoles as inhibitors of 17α-hydroxylase: 17,20-lyase (P450 17): Species and tissue differences*, Pharm. Sci. 3 (1997), pp. 259–263.

[38] J.K. Yano, F. Blasco, H. Li, R.D. Schmid, A. Henne, and T.L. Poulos, *Preliminary characterization and crystal structure of a thermostable cytochrome P450 from Thermus thermophilus*, J. Biol. Chem. 278 (1) (2003), pp. 608–616.

[39] B. Zhao, L. Lei, D.G. Vassylyev, X. Lin, D.E. Cane, and M.R. Waterman, *A dual-function Cytochrome P450 170A1 from* Streptomyces coelicolor. Available at http://www.ebi.ac.uk/pdbsum/3dbg (accessed on November 5, 2009).

[40] B.G. Wachall, M. Hector, Y. Zhuang, and R.W. Hartmann, *Imidazole substituted biphenyls: A new class of highly potent and in vivo active inhibitors of P450 17 as potential therapeutics for treatment of prostate cancer*, Bioorg. Med. Chem. 7 (1999), pp. 1913–1924.

[41] C. Hansch and T. Fujita, $\rho\ \sigma\ \pi$ *analysis: A method for the correlation of biological activity and chemical structure*, J. Am. Chem. Soc. 86 (1964), pp. 1616–1626.

[42] H. Kubinyi, in *Methods and Principles in Medicinal Chemistry*, M.E. Wolff, ed., Vol. 1, VCH, Weinheim, 1993, pp. 497–571.

[43] C. Hansch, A. Leo, and D. Hoekman, *Exploring QSAR. Hydrophobic, Electronic and Steric Constants*, American Chemical Society, Washington, DC, 1995.

[44] *Cerius2 version 4.10*, A product of Accelrys, Inc., San Diego, USA; software available at http://www.accelrys.com/cerius2.

[45] J.T. Leonard and K. Roy, *On selection of training and test sets for the development of predictive QSAR models*, QSAR Comb. Sci. 25 (2006), pp. 235–251.

[46] A.J. Hopfinger and J.S. Tokarsi, *Three-dimensional quantitative structure activity relationship analysis*, in *Practical Applications of Computer-Aided Drug Design*, P.S. Charifson, ed., Marcel Dekker, New York, 1997, pp. 105–164.
[47] A. Hirashima, T. Eiraku, E. Kuwano, and M. Eto, *Three-dimensional molecular field analyses of agonists for tyramine receptor which inhibit sex-pheromone production in* Plodia interpunctella, Internet Electron. J. Mol. Des. 2 (2003), pp. 511–526.
[48] M. Hahn, *Receptor surface models. 1. Definition and construction*, J. Med. Chem. 38 (1995), pp. 2080–2090.
[49] Y. Fan, L.M. Shi, K.W. Kohn, Y. Pommier, and J.N. Weinstein, *Quantitative structure-antitumor activity relationships of camptothecinanalogues: Cluster analysis and genetic algorithm-based studies*, J. Med. Chem. 44 (2001), pp. 3254–3263.
[50] D. Rogers and A.J. Hopfinger, *Application of genetic function approximation to quantitative structure–activity relationship and quantitative structure–property relationship*, J. Chem. Inf. Comput. Sci. 34 (1994), pp. 854–866.
[51] W.J. Dunn III and D. Rogers, *Genetic partial least squares in QSAR*, in *Genetic Algorithms in Molecular Modeling*, J. Devillers, ed., Academic Press, London, 1996, pp. 109–130.
[52] K. Hasegawa, Y. Miyashita, and K. Funatsu, *GA strategy for variable selection in QSAR studies: GA-based PLS analysis of calcium channel antagonists*, J. Chem. Inf. Comput. Sci. 37 (1997), pp. 306–310.
[53] G.W. Snedecor and W.G. Cochran, *Statistical Methods*, Oxford & IBH Publishing Co. Pvt. Ltd, New Delhi, 1967.
[54] S. Wold, *PLS for multivariate linear modeling*, in *Chemometric Methods in Molecular Design*, H. van de Waterbeemd, ed., VCH, Weinheim, 1995, pp. 195–218.
[55] A.K. Debnath, *Quantitative structure–activity relationship (QSAR): A versatile tool in drug design*, in *Combinatorial Library Design and Evaluation: Principles, Software Tools, and Applications in Drug Discovery*, A.K. Ghose and V.N. Viswanadhan, eds., Marcel Dekker, New York, 2001, pp. 73–129.
[56] K. Roy, *On some aspects of validation of predictive QSAR models*, Expert. Opin. Drug. Discov. 2 (2007), pp. 1567–1577.
[57] P.P. Roy and K. Roy, *On some aspects of variable selection for partial least squares regression models*, QSAR Comb. Sci. 27 (2008), pp. 302–313.
[58] K. Roy and P.P. Roy, *Comparative QSAR studies of CYP1A2 inhibitor flavonoids using 2D and 3D descriptors*, Chem. Biol. Drug Des. 72 (2008), pp. 370–382.
[59] P.P. Roy, S. Paul, I. Mitra, and K. Roy, *On two novel parameters for validation of predictive QSAR models*, Molecules 14 (2009), pp. 1660–1701.
[60] K. Roy and S. Paul, *Exploring 2D and 3D QSARs of 2,4-diphenyl-1,3-oxazolines for ovicidal activity against Tetranychus urticae*, QSAR Comb. Sci. 28 (4) (2008), pp. 406–425.
[61] P.P. Roy, J.T. Leonard, and K. Roy, *Exploring the impact of the size of training sets for the development of predictive QSAR models*, Chemom. Intell. Lab. Sys. 90 (2008), pp. 31–42.
[62] L. Eriksson, J. Jaworska, A.P. Worth, M.T. Cronin, R.M. McDowell, and P. Gramatica, *Methods for reliability and uncertainty assessment and for applicability evaluations of classification- and regression-based QSARs*, Environ. Health. Perspect. 111 (2003), pp. 1361–1375.