Journal of Enzyme Inhibition and Medicinal Chemistry, 2010; 25(3): 354–369

informa
healthcare

RESEARCH ARTICLE

Exploring QSAR for CYP11B2 binding affinity and CYP11B2/CYP11B1 selectivity of diverse functional compounds using GFA and G/PLS techniques

Partha P. Roy, and Kunal Roy

Pharmaceutical Technology, Jadavpur University, Kolkata, India

Abstract

A data set of a series of 132 structurally diverse compounds with cytochrome 11B2 and 11B1 (CYP11B2 and CYP11B1) enzyme inhibitory activities was subjected to molecular shape analysis to explore contributions of shape features as well as electronic, structural, and physicochemical parameters toward enzyme inhibitory activities, in search of appropriate molecular scaffolds with optimum substitutions for highly potent CYP11B2 inhibitors. Genetic function approximation (GFA) and genetic partial least squares (G/PLS) were used as chemometric tools for modeling, and the derived equations were of acceptable statistical quality considering both internal and external validation parameters (Q^2: 0.514–0.659, R^2_{pred}: 0.510–0.734). The G/PLS models with spline option for CYP11B2 and CYP11B1 inhibition and selectivity modeling appeared to be the best models based on $r_{m\ (overall)}^2$ criterion. The study indicates the importance of the pyridinylnaphthalene and pyridylmethylene-indane scaffolds with less polar and electrophilic substituents for optimum CYP11B2 inhibitory activity and CYP11B2/CYP11B1 selectivity.

Keywords: *QSAR; CYP11B2; GFA; G/PLS; selectivity*

Introduction

Glucocorticoids and mineralocorticoids in humans are synthesized from cholesterol in different zones of the adrenal cortex[1]. The most potent mineralocorticoid, aldosterone, is located mainly in the adrenal cortex and to a lesser extent in the heart, brain, and vascular smooth muscle cells[2]. It is mainly secreted by the zona glomerulosa of the adrenal gland, and to a minor extent synthesized in the cardiovascular system[3,4]. Adrenal aldosterone synthesis is regulated by the renin–angiotensin–aldosterone system (RAAS) and plasma potassium concentration. The RAAS is an important regulator of blood pressure, and molecular variants in genes that encode components of this system have been associated with several cardiovascular diseases, such as essential hypertension, myocardial infarction, and hypertrophic cardiomyopathy[5,6]. Aldosterone synthase (CYP11B2), a mitochondrial cytochrome P450 enzyme, catalyzes hydroxylation of 11-deoxycorticosterone to corticosterone, and in the next steps it catalyzes the hydroxylation and oxidation in the 18-position of the steroid leading to aldosterone[2] in the adrenal cortex. A related enzyme, 11β-hydroxylase (CYP11B1), is responsible mainly for cortisol biosynthesis, although changes in its activity can influence the biosynthesis of steroid metabolites with mineralocorticoid actions[7]. In humans, CYP11B2 and CYP11B1 proteins are encoded by tandemly duplicated genes on chromosome 8q21–22[8,9], each comprising nine exons whose total nucleotide sequence shares 95% identity[10,11]. To date, three common genetic variants of the aldosterone synthase gene (CYP11B2) have been identified as possible determinants of high blood pressure in patients with essential hypertension[12–14].

Recent studies indicate that activation of the brain RAAS contributes markedly to sympathetic hyperactivity[15–18] as well as left ventricle dysfunction and remodeling after myocardial infarction[19–21]. Treatment with angiotensin converting enzyme (ACE) inhibitors in hypertension and congestive heart failure is found to only transiently reduce circulating aldosterone[22–29], but "aldosterone escape" is observed subsequently, illustrating persistent aldosterone production.

Address for Correspondence: Dr Kunal Roy PhD, Pharmaceutical Technology, Jadavpur University, Kolkata, India. E-mail: kunalroy_in@yahoo.com

(Received 31 March 2009; revised 30 May 2009; accepted 01 June 2009)

ISSN 1475-6366 print/ISSN 1475-6374 online © 2010 Informa UK Ltd
DOI: 10.3109/14756360903179476

Plasma aldosterone concentrations increase even further during long-term treatment with mineralocorticoid receptor (MR) antagonists (e.g. spironolactone)[30,31], which might limit the magnitude of the MR antagonist's protective effect. Moreover, now there is evidence that several aldosterone-induced effects in the cardiovascular system are insensitive to MR antagonists[32].

Promising alternatives to spironolactone and angiotensin-II antagonists[33,34] could be inhibitors of CYP11B2 for the treatment of congestive heart failure and hyperaldosteronism and myocardial fibrosis. Non-steroidal inhibitors should be preferred, for fewer side effects on the endocrine system, to steroidal compounds. Inhibitors of CYP11B2 should not affect 11β-hydroxylase (CYP11B1), which is essential for cortisol formation, to avoid unwanted effects. This aim is certainly not easy to reach, as both enzymes are very similar—there is an identity of 93% at the gene level[35].

Consequent structural optimization of a hit discovered by compound library screening led to a series of nonsteroidal aldosterone synthase inhibitors with high selectivity versus other cytochrome P450 enzymes[36,37]. Different research groups over the years have synthesized and biologically evaluated different classes of compounds to find potential compounds having maximum activity toward the CYP11B2 enzyme[36–42]. However, the selectivity issue has become critical for the design of CYP11B2 inhibitors.

In the present communication we have developed quantitative structure–activity relationship (QSAR) models of different classes of compounds with CYP11B2 and CYP11B1 inhibitory activities reported in the literature[36–41,43], to explore the contribution of different physicochemical, structural, and shape parameters to the inhibitory activity of the compounds toward CYP11B2 and CYP11B1. In addition, we have also developed selectivity QSAR models for CYP11B2 over CYP11B1 enzymes to explore the molecular scaffold and substitutional requirements for optimum selectivity.

Materials and methods

Dataset and descriptors

Inhibitory activities of different classes of compounds toward human CYP11B2 and CYP11B1 enzymes reported in the literature[36–41,43] have been used as the model data set for the present study (Tables 1 and 2). The experimental protocols for determination of enzyme inhibitory activities for all compounds were the same. The inhibitory potencies of the compounds (IC_{50} (nM)) have been converted to the logarithmic scale (pIC_{50} (mM)) and then used for subsequent QSAR analyses as the response variable. The analyses were performed using spatial (radius of gyration, Jurs descriptors, shadow indices, area, PMI-mag, density, Vm), shape (DiFFV, Fo, NCOSV, COSV, shape RMS), thermodynamic (AlogP, AlogP98, Molref), and structural (MW, hydrogen bond donor, hydrogen bond acceptor, chiral centers, number of rotatable bonds) descriptors. For the calculation of 3D descriptors, multiple conformations of each molecule were generated using the optimal search as conformational search method. Each conformer was subjected to an energy minimization procedure using a smart minimizer under open force field (OFF) to generate the lowest energy conformation for each structure. Charges were calculated according to the Gasteiger method. All of the descriptors were calculated using the Descriptor+ module of Cerius2 version 4.10 software running on a Silicon Graphics workstation[44]. Definitions of all descriptors can be found at the Cerius2 tutorial available at the website: http://www.accelrys.com.

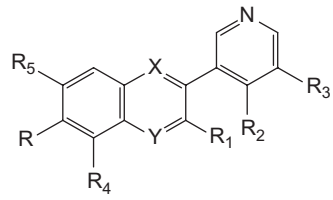
Model development

For the data set under consideration, quantitative CYP11B2 inhibitory activity of all 132 compounds was available while CYP11B1 inhibitory activity was available for 109 compounds. To begin the model development process, the data set ($n = 132$) was classified into clusters using *K*-means clustering based on a standardized spatial, thermodynamic, and structural descriptor matrix (values 0 to 1)[45]. The numbers of compounds for the training set were 99 and 82 for CYP 11B2 and CYP11B1, respectively, while the test sets were composed of 33 and 27 compounds for CYP 11B2 and CYP 11B1, respectively. In addition to the development of respective QSAR models, we also developed selectivity models taking the difference of activity values for two isoenzymes ($\Delta pIC_{50} = pIC_{50\,(CYP11B2)} - pIC_{50\,(CYP11B1)}$) as the response variable. The chemometric tools used for QSAR model development were GFA (genetic function approximation) and G/PLS (genetic partial least squares).

For the computation of shape analysis descriptors, the major steps are: (1) generation of conformers and energy minimization; (2) hypothesizing 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 (molecular shape analysis) descriptors; (6) determination of other molecular features by calculating spatial, electronic, and conformational parameters; (7) selection of conformers; and (8) generation of QSAR equations by the genetic function algorithm (GFA). Optimal search was used as the conformational search method. Each conformer was subjected to an energy minimization procedure using a smart minimizer under open force field (OFF) to generate the lowest energy conformation for each structure. The conformer of the most active compound (compound **24** for both CYP11B1 and CYP11B2) was selected as a shape reference to which all the structures in the study compounds were aligned through pairwise superpositioning. The method used for performing the alignment was a maximum common subgroup (MCSG)[44,46]. This method looks at molecules as points and lines and uses the techniques of graph theory to identify patterns. It finds the largest subset of atoms in the shape

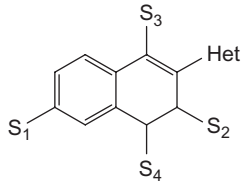
Table 1. Structural features of the diverse compounds having CYP11B2 and CYP11B1 inhibitory activities.

A (Pyridylnaphthalenes)



	X	Y	R	R1	R2	R3	R4	R5
1	CH	CH	-OCH_3	H	H	H		
2	CH	CH	-OCH_3	CH_3	H	H		
3	CH	CH	-CN	H	H	H		
7	CH	CH	-OCH_3	H	-CH_3	H		
8	CH	CH	-CN	H	-CH_3	H		
9	CH	CH	-OCH_3	H	-NH_2	H		
10[a]	CH	CH	-OCH_3	H	H	-OH		
11[a]	CH	CH	-OCH_3	H	H	-OCH_3		
12	CH	CH	-OCH_3	CH_3	H	-OCH_3		
14	CH	CH	-OCH_3	H	H	-OC_2H_5		
15	CH	CH	-OCH_3	H	H	-COOMe		
16[a]	CH	CH	-OCH_3	H	H	-COMe		
17[a]	CH	CH	-OCH_3	H	H	-CH_2COOH		
18	CH	CH	-OCH_3	H	H	-CH_2COOMe		
19	CH	CH	-OCH_3	H	H	-CH_2OH		
20[a]	CH	CH	-OCH_3	H	H	-CH_2OMe		
21	CH	CH	-OCH_3	H	-CH_2OH	H		
22	CH	CH	-OCH_3	H	-CH_2OMe	H		
23	CH	CH	-OCH_3	H	H	-CH(OH)Me		
24	CH	CH	-OCH_3	H	H	-CH(OMe)Me		
25	CH	CH	-OCH_3	H	H	-Ph		
29	CH	CH	H	Benzyl	H	H		
30	CH	CH	-OCH_3	Benzyl	H	H		
31	CH	CH	H	4-Methoxybenzyl	H	H		
32[a]	CH	CH	H	4-Cyanobenzyl	H	H		
33	CH	CH	H	4-Trifluoromethoxybenzyl	H	H		
34[a]	CH	CH	-OCH_3	4-Methoxybenzyl	H	H		
35	CH	CH	H	4-Methoxybenzyl	H	-OCH_3		
36	CH	CH	-OCH_3	4-Methoxybenzyl	H	-OCH_3		
37	CH	CH	-OCH_3	Benzyl	H	-OCH_3		
40	CH	CH	H	4-Methoxybenzoyl	H	H		
41	CH	CH	H	3-Fluoro-4-methoxybenzoyl	H	H		
42[a]	CH	CH	H	4-Cyanobenzoyl	H	H		
43	CH	CH	H	4-Trifluoromethoxybenzoyl	H	H		
44[a]	CH	CH	H	H	H	H		
51	CH	CH	-OH	H	H	H		
52[a]	CH	CH	-Br	H	H	H		
53	CH	CH	-OCH_3	H	H	H	-Cl	H
54[a]	CH	CH	-OCH_3	H	H	H	-Br	H
55	-CHCl	CH	H	H	H	H	H	-OCH_3

B (Dihydropyridylnaphthalenes)



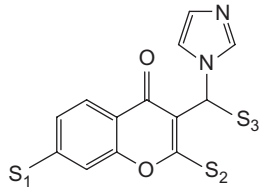
	S1	S2	S3	S4	Het
4[a]	-OCH_3	H	H	H	3-Pyridinyl
5	-OCH_3		H	H	3-Pyridinyl

Table 1. continued on next page.

Table 1. Continued.

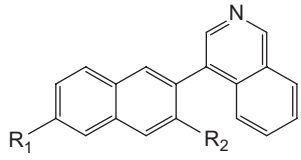
	S1	S2	S3	S4	Het
6[a]	-CN	H	H	H	3-Pyridinyl
13	$-OCH_3$	$-CH_3$	H	H	3-Pyridinyl
28	$-OCH_3$	$-CH_3$	H	H	3-Isoquinolinyl
126	H	H	$-CH_3$	H	3-Pyridinyl
127	H	H	$-C_2H_5$	H	3-Pyridinyl
128[a]	H	$-CH_3$	H	H	3-Pyridinyl
129	H	H	H	$-CH_3$	3-Pyridinyl
130	H	H	H	$-C_2H_5$	3-Pyridinyl

C (Chroman-4-ones)



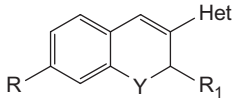
	S1	S2	S3
45	H	Phenyl	
46	H	4-Nitrophenyl	
47	H	4-Bromophenyl	
48	H	4-Methoxyphenyl	
49	$-OCH_3$	Phenyl	
50	H	H	4-Nitrophenyl

D (Naphthaleneisoquinolines)



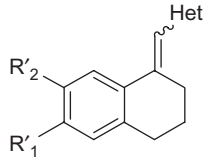
	R1	R2
27	$-OCH_3$	H
28	$-OCH_3$	$-CH_3$
38	H	4-Methoxybenzyl
39[a]	$-OCH_3$	4-Methoxybenzyl

E (Dihydronaphthalenes)



	R	R1	Y	Het
56[a]	H	$-OCH_3$	CH	1-Imidazolyl
57	H	H	CH	5-Imidazolyl
58	H	H	CH	5-Oxazolyl

F (Heteroarylmethylenetetrahydronaphthalenes)



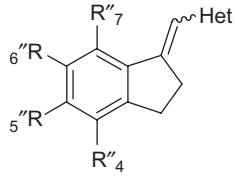
	R'1	R'2	Het	Isomerism
59	H	H	1-H-Imidazolyl	E
60[a]	H	H	1-H-Imidazolyl	Z
63	-CN	H	1-H-Imidazolyl	Z
66[a]	H	-Cl	1-H-Imidazolyl	E
75	H	H	3-Pyridinyl	E
76	H	H	3-Pyridinyl	E
85	$-OCH_3$	H	3-Pyridinyl	E
92[a]	H	H	4-Pyridinyl	E
101	$-OCH_3$	H	4-Pyridinyl	E

Table 1. continued on next page.

Table 1. Continued.

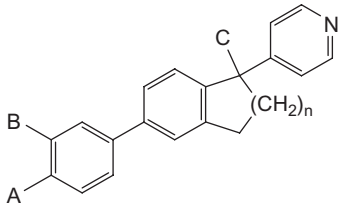
	R′1	R′2	Het	Isomerism
102[a]	-OCH_3	H	4-Pyridinyl	E
105[a]	-OCH_3	-OCH_3	4-Pyridinyl	E

G (Heteroarylmethyleneindanes)



	R″4	R″5	R″6	R″7	Het	Isomerism
61	H	H	H	H	5-Imidazolyl	E
62	H	H	H	H	5-Imidazolyl	Z
64	H	-CN	H	H	5-Imidazolyl	E
65	H	-CN	H	H	5-Imidazolyl	E
67[a]	H	-F	H	H	5-Imidazolyl	E
68	H	-F	H	H	5-Imidazolyl	Z
69[a]	H	-Cl	H	H	5-Imidazolyl	E
70	H	-Cl	H	H	5-Imidazolyl	Z
71	H	-Br	H	H	5-Imidazolyl	E
72	H	-Br	H	H	5-Imidazolyl	Z
73	H	H	H	H	3-Pyridinyl	E
74[a]	H	H	H	H	3-Pyridinyl	Z
77	H	-F	H	H	3-Pyridinyl	E
78[a]	H	-F	H	H	3-Pyridinyl	Z
79[a]	H	-Cl	H	H	3-Pyridinyl	E
80	H	-Cl	H	H	3-Pyridinyl	Z
81	H	-Br	H	H	3-Pyridinyl	E
82	H	-Br	H	H	3-Pyridinyl	Z
83[a]	H	-OCH_3	H	H	3-Pyridinyl	E
84[a]	H	-OCH_3	H	H	3-Pyridinyl	Z
86	-CH_3	H	H	H	3-Pyridinyl	E
87	-F	H	H	H	3-Pyridinyl	E
88	-Cl	H	H	H	3-Pyridinyl	Z
89	-Cl	H	H	H	3-Pyridinyl	E
90	H	H	H	-OCH_3	3-Pyridinyl	E
91	H	H	H	H	4-Pyridinyl	Z
93	H	-F	H	H	4-Pyridinyl	E
94	H	-F	H	H	4-Pyridinyl	Z
95	H	-Cl	H	H	4-Pyridinyl	E
96	H	-Cl	H	H	4-Pyridinyl	Z
97	H	-Br	H	H	4-Pyridinyl	E
98	H	-Br	H	H	4-Pyridinyl	Z
99	H	-OCH_3	H	H	4-Pyridinyl	E
100	H	-OCH_3	H	H	4-Pyridinyl	Z
103	H	H	-OCH_3	H	4-Pyridinyl	E
104	H	H	-OCH_3	H	4-Pyridinyl	Z
106[a]	H	H	-CH_3	H	4-Pyridinyl	E
107	H	H	-CH_3	H	4-Pyridinyl	Z

H (Abiraterones)



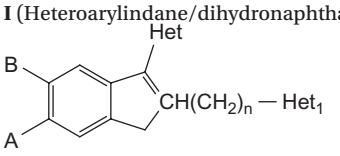
	A	B	C	*n*
108	-OCH_3	-F	-OH	1
109	-F	H	-OH	2

Table 1. continued on next page.

Table 1. Continued.

	A	B	C	*n*
110	-F	-F	-OH	2
111	-OCH_3	-F	-OH	2
113[a]	-F	H	H	1
114[a]	-OH	-OH	H	2
115	-OH	-F	H	2
116[a]	-F	-H	-OH	1
120	-F	H	H	2
121	-F	F	H	2

I (Heteroarylindane/dihydronaphthalenes)



	A	B	Het	Het_1	n
112	4-Fluoro-5-hydroxyphenyl	H	4-Pyridinyl		1
117[a]	4-Fluorophenyl	H	4-Pyridinyl		1
118	4,5-Difluorophenyl	H	4-Pyridinyl		1
119	4,5-Dihydroxyphenyl	H	4-Pyridinyl		1
122	H	H		3-Pyridinyl	0
123	H	H		3-Pyridinyl	1
124	-OCH_3	H		3-Pyridinyl	0
125	-OCH_3	H		3-Pyridinyl	1
131	H	H		1-Imidazoyl	1
132	H	-OCH_3		3-Pyridinyl	0

[a]Test set members.

reference compound that is shared by all the structures in the study table and uses this subset for alignment. A rigid fit of atom pairings was performed to superimpose each structure so that it overlay the shape reference compound. Finally, additional electronic, spatial, and thermodynamic descriptors were also calculated.

The genetic function approximation (GFA) technique[47,48] was used to generate a population of equations rather than one single equation for correlation between biological activity and physicochemical properties. GFA provides an error measure, called the lack of fit (LOF) score, that automatically penalizes models with too many features. It also inspires the use of splines as a powerful tool for non-linear modeling.

The genetic partial least squares (G/PLS) algorithm[49,50] may be used as an alternative to a GFA calculation. G/PLS is derived from two QSAR calculation methods: GFA and partial least squares (PLS). The G/PLS algorithm uses GFA to select appropriate basis functions to be employed 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 parameters such as squared correlation coefficient (R^2) and variance ratio (F) at specified degrees of freedom (df)[51]. For G/PLS equations, the least-squares error (LSE) was taken as an objective function to select an equation, while lack-of-fit (LOF) was noted for the GFA-derived equations. The generated QSAR equations were validated by leave-one-out cross-validation R^2 (Q^2) and predicted residual sum of squares (PRESS)[52–54] 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}), and squared correlation coefficient between observed and predicted values of the test set compounds with (r^2) and without (r_0^2) intercept. It was previously shown that the use of R^2_{pred} and r^2 might not be sufficient to indicate the external validation characteristics[55]. Thus, an additional parameter $r_{m\ (test)}^2$, defined as:

$$r^2 * (1 - \sqrt{r^2 - r_0^2})$$

which penalizes a model for large differences between observed and predicted values of the test set compounds, was also calculated. Two other variants[56] of the r_m^2 parameter, $r_{m\ (LOO)}^2$ and $r_{m\ (overall)}^2$, were also calculated. The parameter $r_{m\ (overall)}^2$ is based on the prediction of both training (LOO prediction) and test set compounds. It was previously shown[56] that $r_{m\ (LOO)}^2$ and $r_{m\ (test)}^2$ penalize a model more strictly than Q^2 and R^2_{pred}, respectively.

Results and discussion

The membership of compounds in different clusters generated using the *K*-means clustering technique is shown in

Table 2. Observed and calculated CYP11B2 and CYP11B1 inhibitory activities and selectivity (CYP11B2/CYP11B1) of different classes of compounds.

	CYP11B2 inhibitory activity			CYP11B1 inhibitory activity		CYP11B2/CYP11B1 selectivity		
Sl. no.	Obs[a]	Cal[b]	Cal[c]	Obs[d]	Cal[e]	Obs[f]	Cal[g]	Cal[h]
Training set								
1	5.208	5.714	5.624	2.802	2.997	2.406	2.618	2.475
2	5.155	5.101	5.242	2.98	2.984	2.175	2.427	2.482
3	5.538	5.131	4.950	3.161	3.331	2.377	2.066	1.884
5	5.481	5.438	5.576	3.606	3.091	1.875	2.099	2.278
7	6.097	5.375	5.354	3.943	3.003	2.154	2.415	2.493
8	6.222	5.599	5.418	4.284	3.396	1.938	1.808	1.843
9	4.886	5.034	5.207	2.818	3.615	2.068	1.332	2.189
12	5.42	5.425	5.832	3.058	3.389	2.362	1.908	2.159
13	5.921	5.264	5.552	4	3.419	1.921	1.765	1.704
14	5.292	5.692	5.557	3.428	3.535	1.864	1.773	2.082
15	6.097	5.789	5.754	4.824	4.236	1.273	1.711	1.729
18	5.161	5.199	4.980	3.701	3.203	1.46	1.379	1.722
19	5.041	4.924	5.198	3.212	2.985	1.829	1.472	1.228
21	4.658	4.736	5.007	2.754	3.164	1.904	1.379	1.310
22	5.658	5.365	5.515	3.362	3.724	2.296	2.043	2.074
23	6.301	5.832	6.406	4.004	4.203	2.297	1.091	1.130
24	6.699	6.009	5.646	5	5.179	1.699	1.405	1.881
25	5.319	5.411	5.449	3.821	3.375	1.498	1.474	1.788
26	6.222	5.417	5.406	4.174	3.318	2.048	1.984	2.163
27	5.509	5.126	5.329	3.074	3.331	2.435	1.837	2.166
28	6.301	4.946	5.020	4.194	3.389	2.107	1.786	2.256
29	3.812	4.163	4.335	3.021	2.775	0.791	2.324	1.776
30	4.276	4.936	4.925	3.194	3.491	1.082	1.593	1.911
31	5.108	4.568	4.808	2.552	2.862	2.556	2.000	1.909
33	5.409	4.995	4.703	2.449	1.888	2.96	2.663	1.998
35	5.114	4.477	4.811	2.742	2.320	2.372	2.329	2.171
36	5.119	4.348	4.723	2.61	2.403	2.509	1.890	2.801
37	4.62	4.674	4.832	2.532	3.024	2.088	1.658	2.108
38	5.523	4.782	4.734	3.105	3.139	2.418	1.689	1.672
40	3.924	4.418	4.987	1.62	2.509	2.304	2.676	1.963
41	4.187	4.176	4.143	1.703	2.234	2.484	2.528	2.044
43	4.553	4.831	4.748	1.947	2.188	2.606	2.451	2.275
45	4.553	4.770	4.574	–	–	–	–	–
46	4.022	3.664	3.242	–	–	–	–	–
47	4.602	4.428	4.595	–	–	–	–	–
48	4.959	4.801	4.740	–	–	–	–	–
49	3.907	5.098	4.759	–	–	–	–	–
50	3.728	3.535	3.473	–	–	–	–	–
51	4.638	4.021	4.390	2.573	2.847	2.065	1.637	0.949
53	4.886	5.506	4.880	2.599	3.148	2.287	2.256	2.136
55	4.538	4.848	4.229	2.565	2.899	1.973	2.457	2.352
57	3.529	4.078	3.933	3.684	3.965	–0.155	0.339	0.259
58	4.921	4.258	4.177	3.094	2.674	1.827	1.692	1.348
59	4.606	4.276	4.445	4.503	3.887	0.103	0.089	–0.040
61	4.387	4.588	4.546	4.587	3.763	–0.2	0.287	0.029
62	4.959	4.740	4.581	5.215	4.916	–0.256	–0.412	–0.553
63	4.644	4.826	4.586	5.161	4.869	–0.517	–0.288	–0.111
64	4.445	4.738	4.538	4.824	3.736	–0.379	0.379	0.373
65	4.447	5.018	4.694	4.91	4.765	–0.463	–0.364	–0.101
68	4.857	3.992	4.166	4.952	5.097	–0.095	–0.470	–0.454
70	5.432	4.928	4.603	4.71	4.740	0.722	–0.554	–0.261
71	4.032	4.342	4.476	4.582	3.441	–0.55	0.111	0.185
72	4.987	4.617	4.621	4.629	4.622	0.358	–0.551	–0.270
73	4.959	4.744	4.542	3.052	3.031	1.907	1.737	1.362
75	4.658	4.204	4.297	3.146	3.230	1.512	1.593	1.290

Table 2. continued on next page.

Table 2. Continued.

Sl. no.	CYP11B2 inhibitory activity			CYP11B1 inhibitory activity		CYP11B2/CYP11B1 selectivity		
	Obs[a]	Cal[b]	Cal[c]	Obs[d]	Cal[e]	Obs[f]	Cal[g]	Cal[h]
76	3.851	4.233	4.096	2.846	3.763	1.005	1.162	1.120
77	5.155	4.049	4.507	3.507	3.277	1.648	1.624	1.435
80	4.137	4.451	4.087	3.569	3.345	0.568	1.127	1.353
81	4.432	4.182	4.275	2.713	2.853	1.719	1.588	1.674
82	3.767	4.107	4.102	3.495	3.217	0.272	1.292	1.347
85	4.244	4.784	5.014	3.044	3.081	1.2	2.105	1.766
86	3.967	4.410	4.077	3.117	3.115	0.85	1.953	1.662
87	4.678	4.224	4.499	3.111	3.278	1.567	1.652	1.617
88	5.046	5.471	4.552	3.517	2.963	1.529	1.633	1.684
89	4.509	5.101	4.575	3.182	3.358	1.327	1.174	1.332
90	4.569	4.770	4.411	3.02	3.601	1.549	1.509	1.115
91	4.745	4.744	4.665	3.031	3.645	1.714	1.173	1.025
93	4.824	5.140	5.674	2.959	3.255	1.865	1.656	1.644
94	4.538	4.264	4.669	4.469	3.857	0.069	1.041	1.131
95	4.745	5.357	5.373	2.82	2.929	1.925	1.628	1.645
96	4.444	4.741	4.435	3.521	3.470	0.923	1.007	1.283
97	4.62	5.043	5.381	2.578	2.825	2.042	1.613	1.643
98	4.252	4.428	4.399	3.315	3.340	0.937	0.993	1.278
99	4.229	5.152	5.090	–	–	–	–	–
100	4.26	4.550	4.485	–	–	–	–	–
101	4.658	5.158	5.334	–	–	–	–	–
103	5.523	5.789	5.719	–	–	–	–	–
104	4.585	4.975	4.813	–	–	–	–	–
107	5.097	4.395	4.403	–	–	–	–	–
108	3.361	3.567	3.396	3.536	2.836	–0.175	0.436	0.219
109	3.076	3.395	3.747	2.936	3.052	0.14	0.742	0.293
110	2.776	3.285	3.078	2.507	2.759	0.269	0.607	0.711
111	3.025	3.445	3.152	3.164	3.046	–0.139	0.375	0.516
112	3.004	3.389	3.518	2.561	2.937	0.443	1.076	0.976
115	3.231	3.328	3.486	2.863	3.009	0.368	0.926	0.443
118	3.507	3.379	3.656	–	–		–	–
119	2.826	3.445	3.464	–	–		–	–
120	3.286	3.746	4.311	–	–		–	–
121	3.246	3.297	3.598	–	–		–	–
122	4.886	4.504	4.634	2.621	2.798	2.265	1.965	1.966
123	5.155	4.784	4.651	2.762	3.032	2.393	1.837	1.812
124	5.398	4.896	4.753	2.245	2.731	3.153	2.674	3.041
125	5.669	5.580	5.453	3.238	2.917	2.431	2.495	2.471
126	5.155	4.618	4.420	2.897	3.115	2.258	1.502	1.282
127	4.523	4.277	4.427	2.674	3.287	1.849	1.367	1.316
129	4.886	4.492	4.613	2.889	3.194	1.997	1.719	1.814
130	3.754	4.078	4.062	2.792	3.372	0.962	1.457	1.772
131	3.476	5.166	5.075	3.194	4.563	0.282	0.897	0.820
132	4.347	5.154	4.937	–	–		–	–
Test set								
4	5.678	5.350	5.318	3.238	2.917	2.44	2.426	2.456
6	5.347	5.022	4.879	3.336	3.345	2.011	1.813	1.849
10	4.027	4.425	4.432	2.049	3.158	1.978	1.889	1.316
11	5.377	5.720	6.092	3.623	3.439	1.754	2.132	2.194
16	5.678	5.251	5.660	3.593	3.009	2.085	2.009	2.014
17	2.915	4.438	3.986	1.423	2.902	1.492	0.824	0.839
20	6.699	5.510	5.238	4.509	3.286	2.19	1.768	2.022
32	5.569	4.635	4.827	2.709	2.145	2.86	2.818	1.688
34	4.959	4.635	4.851	2.364	2.756	2.595	1.719	2.225
39	5.301	4.631	4.683	3.134	3.193	2.167	1.266	2.354
42	4.523	4.184	4.588	2.016	2.418	2.507	2.932	1.800

Table 2. continued on next page.

Table 2. Continued.

	CYP11B2 inhibitory activity			CYP11B1 inhibitory activity		CYP11B2/CYP11B1 selectivity		
Sl. no.	Obs[a]	Cal[b]	Cal[c]	Obs[d]	Cal[e]	Obs[f]	Cal[g]	Cal[h]
44	4.553	4.967	4.886	–	–		–	–
52	4.824	4.668	4.753	2.532	2.906	2.292	1.903	1.708
54	4.481	5.384	5.024	2.349	3.059	2.132	2.190	2.087
56	4.721	4.812	4.502	4.092	4.262	0.629	1.197	1.236
60	5.018	4.811	4.612	5.481	4.966	–0.463	–0.491	–0.493
66	4.325	4.649	4.464	4.728	3.739	–0.403	–0.020	0.181
67	4.777	3.701	3.925	4.686	3.964	0.091	0.224	0.101
69	4.052	4.624	4.536	4.542	3.571	–0.49	0.166	0.201
74	4.036	4.560	4.366	4.06	3.518	–0.024	1.285	0.975
78	4.959	4.313	4.497	3.903	3.728	1.056	1.171	1.052
79	4.585	4.484	4.264	2.832	2.956	1.753	1.602	1.674
83	4.469	4.989	4.957	2.839	2.917	1.63	2.359	1.865
84	4.585	4.351	4.124	3.102	2.917	1.483	1.787	1.345
92	4.824	4.860	5.075	3.845	3.207	0.979	1.489	1.394
102	4.252	4.502	4.646	–	–	–	–	–
105	4.959	5.177	4.774	–	–	–	–	–
106	5.398	5.137	5.085	–	–	–	–	–
113	3.265	3.901	4.322	2.968	3.722	0.297	1.966	1.540
114	2.634	2.966	2.785	2.671	3.083	–0.037	0.735	–0.705
116	3.088	2.990	3.159	–	–		–	–
117	3.917	3.542	4.131	–	–		–	–
128	5.301	4.621	4.552	3.298	3.194	2.003	1.508	1.644

Note. Obs[a], observed CYP11B2 inhibitory activity (refs. 36–41, 43); Cal[b], calculated from Equation (1); Cal[c], calculated from Equation (2); Obs[d], observed CYP11B1 inhibitory activity (refs. 36–41, 43); Cal[e], calculated from Equation (3); Obs[f], observed selectivity (CYP11B2 vs. CYP11B1); Cal[g], calculated from Equation (4); Cal[h], calculated from Equation (5).

Table 3. The test set size was set to approximately 25% of the total data set size[57] 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 Q^2 and R^2_{pred} values greater than 0.5. Models with $r^2_{\text{m (overall)}}$ greater than 0.5 are listed and described in the text. Aligned view of the training set molecules is shown in Figure 1.

Modeling CYP11B2 inhibitory activity

Both linear and linear spline terms were used for development of the models. Variables involved in the GFA-derived models and statistical qualities of the models are shown in Table 4. Equations (1) and (2) are among the best ones obtained from the G/PLS (5000 iterations) method.

$$\begin{aligned} pIC_{50(CYP11B2)} &= 0.406 - 0.016\,Jurs\_TPSA + 0.865\,RadOfGyration \\ &\quad + 0.347\,Hbondacceptor - 0.001\,PMI\_mag \\ &\quad - 0.679\,ShapeRMS + 3.547\,Shadow\_XYfrac + 0.322\,Sr \end{aligned} \quad (1)$$

$$n_{Training} = 99,\ LSE = 0.252,\ R^2 = 0.619,\ F = 30.19(df\ 5,93),$$

$$Q^2 = 0.550,\ PRESS = 30.057,\ n_{Test} = 33,\ R^2_{pred} = 0.536,$$

$$r^2_{m(test)} = 0.529,\ r^2_{m(LOO)} = 0.528,\ r^2_{m(overall)} = 0.527$$

According to the standardized regression coefficients, the relative order of importance of the descriptors is the following: Jurs_TPSA > RadOfGyration > PMI_mag > Hbondacceptor > ShapeRMS > Shadow_XYfrac > Sr.

Jurs_TPSA has negative impact on the inhibitory potency toward the CYP11B2 enzyme. Total polar surface area (Jurs_TPSA) is the sum of solvent-accessible surface areas of atoms with absolute values of partial charges greater than 0.2, and can be expressed by the following formula:

$$TPSA = \sum\nolimits_a SA_a$$

$$|q_a| \geq 0.2$$

Compounds having a lower value of this parameter have higher inhibitory activity. The presence of a lower number of polar groups or fragments (e.g. compounds **89** and **127** of pyridylmethylene-indane and pyridinylnaphthalene scaffolds, respectively, with few or lower number of polar substitutions) increases the CYP11B2 inhibitory activity. Compounds with a high number of polar groups or fragments (abiraterone analogs, e.g. **109**, **110**, and **111** having F and hydroxyl groups) have poor inhibitory activity. Compounds with a pyridinylnaphthalene scaffold (e.g. **3** and **127**) and pyridylmethylene-indane scaffold (**80**, **82**, **89**, **107**) have higher inhibitory activity than imidazole-substituted chroman-4-one analogs (e.g. compounds **46**, **50**) and abiraterone analogs (e.g. compound **110**).

The radius of gyration is a measure of the size of an object, a surface, or an ensemble of points. It is calculated as the

Table 3. *K*-means clustering of compounds using standardized descriptors.

Cluster no.	No. of compounds in cluster	Compounds (Sl. nos.) in each cluster											
1	5	**1**	**3**	**55**	**56**	**136**							
2	12	**83**	**92**	**93**	**101**	**104**	**112**	**113**	**114**	**115**	**116**	**119**	**139**
3	9	**52**	**65**	**66**	**69**	**72**	**75**	**76**	**77**	**79**			
4	15	**62**	**67**	**68**	**70**	**71**	**73**	**81**	**82**	**84**	**97**	**103**	**118**
		135	**138**	**144**									
5	9	**5**	**13**	**25**	**29**	**125**	**132**	**140**	**141**	**142**			
6	16	**6**	**8**	**45**	**74**	**78**	**80**	**85**	**86**	**88**	**98**	**99**	**100**
		105	**106**	**108**	**134**								
7	9	**2**	**4**	**7**	**9**	**58**	**91**	**111**	**137**	**147**			
8	7	**53**	**87**	**89**	**90**	**107**	**109**	**110**					
9	9	**24**	**120**	**121**	**122**	**123**	**126**	**127**	**128**	**133**			
10	8	**10**	**18**	**20**	**22**	**47**	**51**	**124**	**131**				
11	18	**11**	**12**	**14**	**15**	**17**	**19**	**21**	**23**	**26**	**27**	**28**	**31**
		46	**49**	**50**	**117**	**129**	**130**						
12	15	**30**	**32**	**33**	**34**	**35**	**36**	**37**	**38**	**39**	**40**	**41**	**42**
		43	**44**	**48**									

root mean square distance of the object's parts from either its center of gravity or an axis. This can be calculated by the following equation:

$$RadofGyration = \sqrt{\left(\sum \frac{(x_i^2 + y_i^2 + z_i^2)}{N}\right)}$$

where N is the number of atoms and x, y, z are the atomic coordinates relative to the center of mass. A positive coefficient of RadOfGyration indicates compounds with high values of the parameter (e.g. **15**, **19**, and **26** having a pyridinylnaphthalene nucleus with fewer polar substitutions at the 5' position of the pyridine ring) are more potent inhibitors than compounds (e.g. **62**, **68**, **70**) having low values of the parameter. Although abiraterone analogs (e.g. compounds **108**, **111**, and **119)** have high values of RadOfGyration, these compounds also have high values of Jurs_TPSA and, hence, they show poor inhibitory activity due to the presence of the electronegative fluorine atom as well as the hydroxyl group, which increases the polarity of the molecules. Principal moment of inertia (PMI_mag) has a negative contribution toward the inhibitory activity.

The positive coefficient of Hbondacceptor (hydrogen bond acceptor) indicates that an increase in the number of hydrogen bond acceptor groups increases the inhibitory activity. Compounds **43**, **46**, and **50** (pyridinylnaphthalene and chroman-4-one nucleus with more polar substitution, e.g. -OCF_3, -NO_2) have the maximum number of hydrogen bond acceptor groups, but the activity of these compounds is relatively less because of high Jurs_TPSA values. On the other hand, **29** and **57** have less hydrogen bond acceptor groups and possess less activity.

Root mean square (RMS) deviation between the individual molecule and the shape reference compound (ShapeRMS) also has negative impact toward the inhibitory activity. Compounds **15** and **23** have the minimum root mean square deviation from the shape reference compound **24** and the former compounds possess significantly higher activity than compounds **65** and **123** having maximum ShapeRMS values.

Compounds **12**, **23**, and **24** have high Shadow_XYfrac values and significantly higher inhibitory potency than compounds **40** and **50**. Pyridinylnaphthalene compounds with substitutions such as -OCH_3, -CH(OH)Me, and -CH(OMe)Me at the 5' position of the pyridine ring of the pyridinylnaphthalene nucleus (**12**, **23**, **24**) show maximum values for Shadow_XYfrac.

The positive coefficient of superdelocalizability (Sr) has a positive impact on the inhibitory activity. This indicates that an increase in the electrophilic property is conducive for inhibitory activity. Compounds **55** and **86** having a lower value of the parameter have relatively less activity than compounds **26**, **88**, and **103** having a high value of the parameter.

$$\begin{aligned} pIC_{50(CYP11B2)} &= 6.475 - 0.870\,ShapeRMS - 0.021 < Jurs\_TPSA \\ &\quad - 70.8915 > -0.482 < 1.33355 - Sr > \\ &\quad - 0.018 < 88.2511 - Shadow\_XY > \\ &\quad + 11.202 < Shadow\_XYfrac - 0.651254 > \\ &\quad - 0.101 < 4.18953 - Jurs\_RNCS > -0.095\,A\log P98 \end{aligned} \quad (2)$$

$$n_{Training} = 99,\ LSE = 0.212,\ R^2 = 0.614,\ PRESS = 30.507,$$
$$F = 76.47(df\,2,96),\ Q^2 = 0.543,\ n_{Test} = 33,\ R^2_{pred} = 0.582,$$
$$r^2_{m(test)} = 0.565,\ r^2_{m(LOO)} = 0.522,\ r^2_{m(overall)} = 0.554$$

The relative importance of the descriptors according to their standardized value is in the following order: ShapeRMS > <Jurs_TPSA-70.8915> > <1.33355-Sr> > <88.2511-Shadow_XY> > <Shadow_XYfrac-0.651254> > <4.18953-Jurs_RNCS> > AlogP98.

The term <Jurs_TPSA-70.8915> shows negative contribution toward the inhibitory activity, similar to Jurs_TPSA

Table 4. Comparative table of statistical qualities of different models

	Model	Variables	R^2	Q^2	R^2_{pred}	$r^2_{m\ (test)}$	$r^2_{m\ (LOO)}$	$r^2_{m\ (overall)}$
CYP11B2 modeling								
GFA	M1	Jurs_TPSA, Jurs_FNSA_3, ShapeRMS, RadOfGyration, Jurs_RASA, Jurs_WPSA_2	0.599	0.543	0.641	0.559	0.385	0.432
GFA spline	M2	<Jurs_TPSA-55.2527>, <2.4853-AlogP98>, <LUMO-1.68653>, <ShapeRMS-0.655548>, <COSV-163.577>, Density	0.672	0.630	0.611	0.569	0.472	0.447
G/PLS	M3 Eq. (1)	ShapeRMS, Hbondacceptor, Jurs_TPSA, RadOfGyration, Sr, Shadow_XYfrac, PMI_mag	0.619	0.550	0.536	0.529	0.528	**0.527**[a]
G/PLS spline	M4 Eq. (2)	<1.33555-Sr>, <88.2511-Shadow_XY>, <Shadow_XYfrac-0.651254>, <4.18953-Jurs_RNCS>, <Jurs_TPSA-70.8915>, ShapeRMS, AlogP98	0.614	0.543	0.583	0.565	0.522	**0.554**[a]
CYP11B1 modeling								
GFA	M5	Jurs_WNSA_2, Jurs_RNCS, Shadow_XYfrac, Jurs_DPSA_2, Jurs_RPCG, Jurs_SASA	0.620	0.519	0.502	0.479	0.390	0.409
GFA spline	M6	<Shadow_Ylength-9.01347>, Jurs_RNCS, <COSV-167.814>, <14.47-Shadow_Xlength>	0.641	0.598	0.543	0.511	0.440	0.454
G/PLS	M7	Jurs_RPCG, Jurs_RNCS, ShapeRMS, Shadow_XYfrac, Jurs_WPSA_1, Jurs_WNSA_2, Shadow_nu	0.595	0.514	0.510	0.496	0.489	0.498
G/PLS spline	M8 Eq. (3)	<4.04134- Jurs_RNCS>, <Shadow_YZ-41.056>, <3.51287-RadOfGyration>, <COSV-167.814>, Jurs_RNCG, <-9.59302-Jurs_WNSA_3>	0.645	0.571	0.548	0.488	0.539	**0.558**[a]
Selectivity modeling								
GFA	M9	ShapeRMS, Jurs_RNCS, PMI_mag, Jurs_RPCS, Shadow_XY	0.609	0.545	0.643	0.560	0.521	0.529
GFA spline	M10 Eq. (4)	Jurs_RNCS, <Shadow_Ylength-9.08574>, <1-Hbonddonor>, <4.12361-RadOfGyration>, Jurs_SASA	0.708	0.659	0.653	0.618	0.578	**0.588**[a]
G/PLS	M11	Chiral centers, Jurs_RNCS, Jurs_RPCS, Rotatablebonds, Shadow_nu, Jurs_TPSA, COSV	0.634	0.557	0.734	0.653	0.332	0.452
G/PLS spline	M12 Eq. (5)	<7.40455-Shadow_Ylength>, <LUMO-2.06415>, <6.0479-Jurs_RNCS>, <Jurs_WNSA_2+139.393>, Hbonddonor, <Jurs_SASA-584.135>, <4.12361-RadOfGyration>	0.709	0.634	0.718	0.702	0.604	**0.633**[a]

[a]Only models with $r^2_{m\ (overall)}$ values more than 0.5 (Equations (1)–(5)) have been discussed in detail in the text.

(Equation (1)). For activity of the compounds, the value of Jurs_TPSA should be less than 70.8915. Compounds such as **46**, **50**, **110**, **112**, **115**, and **119** (imidazole-substituted chroman 4-one and abiraterone nucleus) have relatively large values of Jurs_TPSA and significantly poor inhibition potential compared with compounds **1**, **2**, **7**, **8**, and **24** (pyridinylnaphthalene scaffold) having a low value of Jurs_TPSA.

The term <1.33355-Sr> has negative impact, although the term Sr shows a positive contribution in Equation (1) toward inhibitory activity. This indicates that for optimum inhibitory activity, the value of Sr should be greater than 1.33355.

The area of molecular shadow at the XY plane (Shadow_XY) should be greater than 88.2511 for activity, as the term <88.2511-Shadw_XY> shows negative contribution toward the activity. Compounds **14**, **15**, **18**, **23**, **24**, **25**, **28**, **30**, **35**, **36**, and **38** (pyridinylnaphthalene scaffold analogs) having zero value for the term show higher inhibitory potential than compounds such as **57**, **62**, **68**, **91**, and **94** having a non-zero value. Substitutions with groups such as -OEt, -COOMe, -CH_2COOMe, -CH(OH)Me, -CH(OMe)Me, and phenyl at the 5' position of the pyridine ring linked to the naphthalene nucleus increase the value of Shadow_XY and consequently increase the inhibitory activity. The positive coefficient of the term <Shadow_XYfrac-0.651254> indicates that for optimal activity, values of Shadow_XYfrac should be greater than 0.651254 (compounds with pyridinylnaphthalene scaffold, e.g. **12**, **23**, **24** possess significantly higher activity than compounds **40**, **50**).

The term <4.18953-Jurs_RNCS> has negative impact on the inhibitory activity. This indicates that the relative negative charge surface area (Jurs_RNCS) should be greater than 4.18953 for the desired biological activity. Jurs_RNCS is defined as the solvent-accessible surface area of the most negative atom divided by the relative negative charge (RNCG), i.e.:

$$RNCS = \frac{SA^-_{max}}{RNCG}$$

$$RNCG = \frac{Q^-_{max}}{Q^-}$$

where Q^-_{max} is the charge of the most negative atom and Q^- is the total negative charge.

This indicates that if the number of electronegative atoms in a molecule increases then RNCG will decrease and

Figure 1. Aligned geometry of the training set compounds.

Jurs_RNCS will increase. Compounds **1**, **2**, **3**, **5**, **7**, **8**, **12**, **13**, **14**, **26**, and **33** (pyridinylnaphthalene scaffold) show more activity than compounds such as **45**, **47**, **62**, **65**, **68**, and **131**. The negative coefficient of AlogP98 (measure of hydrophobicity) indicates that the more is the lipophilicity, the less is the inhibitory activity.

Modeling CYP11B1 inhibitory activity

Both linear and linear spline terms were used for development of the models. Equation (3) is among the best obtained from the G/PLS (5000 iterations) method. Other models are listed in Table 4 along with their statistical qualities.

$$\begin{aligned} pIC_{50(CYP11B1)} &= 6.215 - 9.616\,Jurs\_RNCG - 0.116 < -9.593 \\ &\quad - Jurs\_WNSA\_3 > -0.057 < Shadow\_YZ \\ &\quad - 41.056 > + 0.301 < 4.04134 - Jurs\_RNCS > \\ &\quad + 2.989 < 3.51287 - RadOfGyration > \\ &\quad + 0.017 < COSV - 167.814 > \end{aligned} \tag{3}$$

$$n_{Training} = 82,\ LSE = 0.198,\ R^2 = 0.645,\ PRESS = 22.964,$$
$$F = 27.60(df\,5,76),\ Q^2 = 0.571,\ n_{Test} = 27,\ R^2_{pred} = 0.548,$$
$$r^2_{m(test)} = 0.488,\ r^2_{m(LOO)} = 0.539,\ r^2_{m(overall)} = 0.558$$

The relative order of importance of the descriptors is as follows: Jurs_RNCG > <-9.59302-Jurs_WNSA_3> > <Shadow_YZ-41.056> > <4.04134-Jurs_RNCS> > <3.51287-RadOfGyration> > <COSV-167.814> (some values have been rounded to three decimal places in Equation (3)).

The negative coefficient of Jurs_RNCG indicates that a low value for this parameter is conducive for activity. However, compounds such as **33**, **36**, **41**, and **43** (pyridinylnaphthalene nucleus) have low values of Jurs_RNCG and these compounds possess poor inhibitory activity. The reason behind this is that the term <-9.59302-Jurs_WNSA_3> has maximum value for these compounds. As the term <-9.59302-Jurs_WNSA_3> has negative contribution toward the activity, this leads to poor inhibitory potency of these compounds. Compounds **66** and **136** have high values of RNCG and low activity. Compounds such as **71** and **72** of imidazolylmethylene-indane scaffold show good inhibitory activity in spite of high RNCG values because of the zero value of the term <-9.59302-Jurs_WNSA_3>. It has been observed that compounds with moderate Jurs_RNCG values (e.g. **8**, **15**, **23**, **61**, **62**, **63**, **68**, **71**, **72**, and **94**) show significant CYP11B1 inhibitory activity.

The negative coefficient of the term <-9.59302-Jurs_WNSA_3> indicates that the value of Jurs_WNSA_3 should be less negative than –9.59302 for optimum activity. Compounds with Jurs_WNSA_3 values more negative than –9.59302 (as in the case of pyridinylnaphthalene (**31**, **33**, **41**, and **43**) and abiraterone (**108**, **110**, and **111**) analogs) have less inhibitory activity than compounds such as **62**, **76**, **91**, **126**, and **131**. The negative coefficient of the term <Shadow_YZ-41.056> indicates that compounds should have an area of molecular shadow in the YZ plane of less than 41.056. Compounds **29**, **31**, **33**, **35**, **36**, **38**, **40**, **41**, and **43** (pyridinylnaphthalene scaffold) have maximum values for the term and poor inhibitory activity. The term <4.04134-Jurs_RNCS> has positive contribution toward the inhibitory activity. This indicates that values of Jurs_RNCS should be less than 4.04134 for optimum activity. Compounds **62**, **63**, **65**, **68**, **70**, and **72** with imidazolylmethylene-indane scaffold have low values of Jurs_RNCS and significant inhibitory activity compared with compounds such as **1**, **2**, **3**, **5**, and **7** (pyridinylnaphthalene scaffold) with higher Jurs_RNCS values. The term <3.51287-RadOfGyration> has a positive contribution toward the inhibitory activity and compounds such as **62**, **63**, **68**, **70**, and **72** have high values for the term and significant inhibitory activity. Compounds **1**, **2**, **5**, **7**, **8**, **9**, and **12** have high values of RadOfGyration and thus negatively contribute to the inhibitory activity.

Common steric overlap volume (COSV) of the compounds should be greater than 167.814 for the inhibitory activity as the term <COSV-167.814> has a positive effect on the inhibitory activity. Compounds **15**, **23**, and **24** have high activity values and **2**, **3**, **9**, and **12** have lower values for the term and low activity, although all the compounds have a pyridinylnaphthalene scaffold. This indicates the importance of substitutions at the 5' position of the pyridine ring of the pyridinylnaphthalene scaffold, which are absent in the case of the second set of compounds.

Selectivity modeling

In addition to the development of respective QSAR models, we have also developed selectivity models, taking the

difference between two activities ($\Delta pIC_{50} = pIC_{50\ (CYP11B2)} - pIC_{50\ (CYP11B1)}$) as the response variable. Both linear and linear spline terms were used for development of the models. Equations (4) and (5) are among the best obtained from genetic function approximation (5000 iterations) and G/PLS (5000 iterations) methods. Other models are listed in Table 4.

$$\begin{aligned}\Delta pIC_{50(CYP11B2-CYP11B1)} &= 5.714(\pm 1.133) - 010(\pm 0.002)\,Jurs\_SASA \\ &\quad - 2.373(\pm 0.329) < 4.12361 - RadOfGyration > \\ &\quad + 0.328(\pm 0.071) < Shadow\_Ylength - 9.08574 > \\ &\quad + 0.085(\pm 0.029)\,Jurs\_RNCS + 0.882(\pm 0.156) < 1 \\ &\quad - Hbonddonor > \end{aligned} \tag{4}$$

$$n_{Training} = 82,\ LOF = 0.374,\ R^2 = 0.708,\ PRESS = 25.483,$$
$$F = 36.93(df\ 5,76),\ Q^2 = 0.658,\ n_{Test} = 27,\ R^2_{pred} = 0.653,$$
$$r^2_{m(test)} = 0.618,\ r^2_{m(LOO)} = 0.578,\ r^2_{m(overall)} = 0.588$$

The relative importance of the descriptors when they are regressed with the standardized values is in the following order: Jurs_SASA > <4.12361-RadOfGyration> > <Shadow_Ylength-9.08574> > Jurs_RNCS > <1-Hbonddonor>.

Solvent-accessible surface area (SASA) is calculated using a sphere of radius 1.5 Å to approximate the contact surface formed when a water molecule interacts with the considered molecule. The negative coefficient of Jurs_SASA indicates that an increase in the solvent-accessible surface area (Jurs_SASA) decreases the selectivity of the compound toward the CYP11B2 enzyme. Compounds **58** and **122** have low values of the parameter and significant selectivity toward the CYP11B2 enzyme. Similarly, abiraterone analogs (**108**, **111**) have high values of the parameter and these are least selective. Compounds such as **57**, **61**, **62**, and **68** have low values of Jurs_SASA and at the same time have a high value of <4.12361-RadOfGyration>, and these are least selective. Again, compounds **35**, **36**, **37**, and **43** (pyridinylnaphthalene scaffold) have high selectivity toward CYP11B2 besides high Jurs_SASA value because of a zero value of the term <4.12361-RadOfGyration> and moderate values of Jurs_RNCS. A negative coefficient of the term <4.12361-RadOfGyration> indicates that selective compounds should have values of RadOfGyration greater than 4.12361. In this case, the same results are obtained as seen in Equations (1) and (2). The length of the molecule in the Y dimension (Shadow_Ylength) should be greater than 9.08574 for selectivity. Compounds **34**, **36**, **37**, **41**, **42**, and **44** (pyridinylnaphthalene nucleus) have maximum value for the parameter and significant selectivity toward the CYP11B2 enzyme. Compounds such as **7**, **8**, **9**, **18** (pyridinylnaphthalene congeners), and **97** (pyrimidylmethylene indane) also have good selectivity because of moderate Jurs_SASA and Jurs_RNCS values. The relative negative surface area (Jurs_RNCS) has positive contribution toward selectivity of the CYP11B2 enzyme. Compounds **51**, **81**, **95** (pyrimidylmethylene indane), and **97** (pyrimidylmethylene indane) have high values and considerably high selectivity compared with compounds such as **57**, **62**, **63**, **65**, **68**, **70**, **72**, and **131**. The term <1-Hbonddonor> has a positive coefficient indicating that the number of hydrogen bond donors should be 1 or 0 for optimal CYP11B2 selectivity. Compounds **1**, **2**, **3**, **5**, **7**, **8**, **12**, **13**, **14**, **15**, **18**, **22**, **24**, **25**, **26**, **27**, **28**, **31**, **33**, **35**, **36**, and **37** (pyridinylnaphthalene scaffold with optimal substitutions at different positions) have zero values for the hydrogen bond donor parameter and significant selectivity toward the CYP11B2 enzyme.

$$\begin{aligned}\Delta pIC_{50(CYP11B2-CYP11B1)} &= 1.857 - 1.666 < 4.12361 - RadOfGyration > \\ &\quad - 0.881 Hbonddonor + 0.009 < Jurs\_WNSA\_2 \\ &\quad + 139.393 > + 13.723 < 7.40455 - Shadow\_Ylength > \\ &\quad - 1.564 < LUMO - 2.06415 > + 0.015 < Jurs\_SASA \\ &\quad - 584.135 > - 0.069 < 6.0479 - Jurs\_RNCS > \end{aligned} \tag{5}$$

$$n_{Training} = 82,\ LSE = 0.254,\ R^2 = 0.709,\ PRESS = 27.245,$$
$$F = 37.13(df\ 5,76),\ Q^2 = 0.634,\ n_{Test} = 27,\ R^2_{pred} = 0.718,$$
$$r^2_{m(test)} = 0.702,\ r^2_{m(LOO)} = 0.604,\ r^2_{m(overall)} = 0.663$$

According to standardized coefficients, the relative importance of the descriptors is in the following order: <4.12361-RadOfGyration> > Hbonddonor > <Jurs_WNSA_2 + 139.393> > <7.40455-Shadow_Ylength> > <LUMO-2.06415> > <Jurs_SASA-584.135> > <6.0479-Jurs_RNCS>.

The term hydrogen bond donor has negative contribution toward the selectivity. This indicates that an increase in the number of hydrogen bond donor groups will decrease the selectivity. Compounds with no hydrogen bond donor group such as **1**, **2**, **3**, **5**, **7**, **8**, **12**, **13**, **14**, **15**, **18**, **22**, **24**, **25**, **26**, **27**, **28**, **31**, **33**, **35**, **36**, and **37** (pyridinylnaphthalene scaffold) have significant selectivity. Again, compounds with value 1 for the hydrogen bond donor term, e.g. **62**, **63**, **64**, **65**, **68**, **70**, **71**, **72** (imidazolylmethylene-indane scaffold) and **108**, **109**, **110**, **111** (abiraterone analogs) have least selectivity because of the high value of the term <4.12361-RadOfGyration> as well as <6.0479-Jurs_RNCS>.

The term <Jurs_WNSA_2 + 139.393> has a positive coefficient indicating that the value of WNSA_2 should be less than 139.393. The term <7.40455-Shadow_Ylength> also has a positive coefficient. Compounds **7**, **9**, **134**, and **136** have high values and significant selectivity. Similar to Equation (4), we find here that the optimum value of the Shadow_Ylength should be less than 7.40. The term <LUMO-2.04615> has a negative coefficient which indicates that values of LUMO should be less than 2.04615 (more electrophilic). Compounds such as **1**, **2**, **3**, **7**, **9**, **12**, **22**, **23**, **26**, **27**, **28**, **31**, **33**, **35**, **36**, **37**, **38**, **40**, **41**, and **43** (pyridinylnaphthalene nucleus) having lower values of LUMO have higher selectivity than compounds with the imidazolemethylene-indane scaffold (**61**, **62**) and abiraterone analogs (**108**, **109**, **111**).

The values of Jurs_SASA should be more than 584.135 for higher selectivity as the term <Jurs_SASA-584.135> has a positive coefficient. Compounds **33**, **35**, **36**, **37**, **41**, and **43**

(pyridinylnaphthalene nucleus) have higher selectivity than compounds such as **65**, **69**, **70**, **73**, **76**, and **79** of imidazolylmethylene-indane and pyridylmethylene-indane scaffolds. The relative negative surface area (Jurs_RNCS) should be greater than 6.0479 for optimum selectivity. In the earlier case it was observed that Jurs_RNCS has a positive impact on selectivity, and here we get a limiting value of Jurs_RNCS for selectivity. Compounds such as **57**, **59**, **61**, **62**, **63**, **64**, **65**, **6**8, **70**, **7**1, **72**, and **131** have low RNCS values and these are least selective. On the other hand, compounds such as **2**, **3**, **8**, **19**, **23**, **51**, **58**, and **73** have high selectivity as they have high Jurs_RNCS values.

Overview

From the discussions on Equations (1) and (2) it can be observed that compounds with pyridinylnaphthalene and pyridylmethylene-indane scaffolds show significant CYP11B2 inhibitory activity. The results also suggest that less polar substitution (-OEt, -COOMe, -CH_2COOMe, -CH(OH)Me, -CH(OMe)Me, phenyl) at the 5′ position of the pyridine ring of the pyridinylnaphthalene series of compounds facilitate the CYP11B2 inhibitory activity, which was supported by a previous article[39]. Isoquinoline compounds also show potent CYP11B2 activity. The introduction of the benzyl residue with substitutions at the 3 position of the naphthalene nucleus shows significant CYP11B2 inhibitory activity. Abiraterone analogs were found to have poor inhibitory activity for CYP11B2. From the discussion on Equation (3), it can be observed that effective inhibitors of CYP11B1 are the pyridinylnaphthalene and imidazolemethylene-indane scaffold congeners. The abiraterone analogs show poor inhibitory activity for CYP11B1. From the discussions on Equations (4) and (5), it is found that selectivity toward human CYP11B2 is higher in the case of pyridinylnaphthalene and to a lesser extent the pyridylmethylene-indane scaffold. Substitutional requirements are the same as mentioned in the case of CYP11B2 binding. Imidazolemethylene-substituted indane compounds and abiraterone analogs are found to be less selective toward CYP11B2. Imidazole-substituted naphthalene, tetrahydronaphthalene, and chroman-4-one nucleus did not contribute to the selectivity.

Conclusion

The whole dataset ($n = 132$ for CYP11B2 and $n = 109$ for CYP11B1) was divided into a training set (99 and 82 compounds for CYP11B2 and CYP11B1, respectively) and a test set (33 and 27 compounds for CYP11B2 and CYP11B1, respectively) based on *K*-means clustering of the standardized spatial, thermodynamic, and structural descriptor matrix, and models were developed from the training set. The predictive ability of the models was judged from the prediction of CYP11B2 and CYP11B1 inhibition activity of the test set compounds. A comparison of statistical quality of different models is given in Table 4. Only those models showing values of $r^2_{m\ (overall)}$ more than 0.5 have been discussed in the text. In all the models, the difference between R^2 and Q^2 is not very high (less than 0.3)[58]. All 12 models have Q^2 and R^2_{pred} values greater than 0.5. For CYP11B2 inhibition, the GFA model with spline option (model M2) was found to be the best ($Q^2 = 0.630$), and the best predictive model was the GFA model with linear option (model M1, $R^2_{pred} = 0.641$). Based on $r^2_{m\ (overall)}$ criteria, the best model among the four (Table 4) was the G/PLS model with spline option (model M4; Equation (2); $r^2_{m\ (overall)} = 0.554$). In the case of CYP11B1 inhibitory activity, the GFA and G/PLS models with spline options were superior to the models derived with linear options. Model M6 (GFA model with spline option) was the best model considering the Q^2 value (0.598), while model M8 (Equation (3); G/PLS model with spline option) was the best predictive model ($R^2_{pred} = 0.548$) as well as the best model according to $r^2_{m\ (overall)}$ criteria (0.558). The GFA model with spline option for selectivity (model M10) was found to be the best model based on internal validation statistics ($Q^2 = 0.659$), and the G/PLS model for selectivity (model M11) was found to be the best predictive model ($R^2_{pred} = 0.734$). However, based on the $r^2_{m\ (overall)}$ criterion, which considers both internal validation and external validation, the G/PLS model with spline option (model M12; Equation (5)) appeared to be the best ($r^2_{m\ (overall)} = 0.633$). It was evident that for CYP11B2 and CYP11B1 inhibition and selectivity modeling, the G/PLS with spline options appeared to be the best models based on the $r^2_{m\ (overall)}$ criterion. Modeling with CYP11B2 inhibitory activity indicates that the distribution of charges and surface area appear to be very important for CYP11B2 binding and selectivity. Electrophilicity is also important for CYP11B2 selectivity. The derived models also indicate the importance of pyridinylnaphthalene and pyridylmethylene-indane scaffolds for optimum inhibitory activity, with appropriate less polar substitutions in appropriate positions. As the CYP11B2 and CYP11B1 binding affinity values are dependent on multiple factors, medicinal chemists should design novel compounds in such a way whereby the factors contributing positively to the CYP11B2 binding affinity and 11B2/11B1 selectivity are enhanced and detrimental factors are reduced. The models developed in this study could be helpful in designing new potent selective CYP11B2 inhibitors.

***Declaration of interest*:** This research work is supported by a Major Research Grant of the University Grant Commission (UGC), New Delhi. One of the authors (P.P.R.) thanks the UGC, New Delhi, for a fellowship.

References

1. Benjamin B, Karsten D, Rita B. Conferring aldosterone synthesis to human CYP11B1 by replacing key amino acid residues with CYP11B2-specific ones. Eur J Biochem 1998;252:458–66.
2. Kawamoto T, Mitsuuchi Y, Toda K, Yokoyama Y, Miyakara K, Miura S, et al. Role of steroid 11beta-hydroxylase and steroid 18-hydroxylase in

the biosynthesis of glucocorticoids and mineralocorticoids in humans. Proc Natl Acad Sci USA 1992;89:1458–62.
3. Takeda Y. Vascular synthesis of aldosterone: role in hypertension. Mol Cell Endocrinol 2004;217:75–9.
4. Davies E, MacKenzie SM. Extra-adrenal production of corticosteroids. Clin Exp Pharmacol Physiol 2003;30:437–45.
5. Brilla CG. Renin-angiotensin-aldosterone system and myocardial fibrosis. Cardiol Vasc Res 2000;47:1–3.
6. Lijnen P, Petrov V. Induction of cardiac fibrosis by aldosterone. J Mol Cell Cardiol 2000;32:865–79.
7. Hlubocká Z, Jáchymová M, Heller S, Umnerová V, Danzig VV, Lánská V, et al. Association of the -344T/C aldosterone synthase gene variant with essential hypertension. Physiol Res 2008 Dec 17. [Epub ahead of print]
8. Chua SC, Szabo P, Vitek A, Grzeschik KH, John M, White PC. Cloning of cDNA encoding steroid 11 beta-hydroxylase (P450c11). Proc Natl Acad Sci USA 1987;84:7193–7.
9. Wagner MJ, Ge Y, Siciliano M, Wells DE. A hybrid phatase using a novel monoclonal antibody and peptide elution, cell mapping panel for regional localization of probes to human chromosome 8. Genomics 1991;10:114–25.
10. Mornet E, Dupont J, Vitek A, White PC. Characterizing two genes encoding human steroid 11beta-hydroxylase (P-450(11) beta). J Biol Chem 1989;264:20961–7.
11. Kawamoto T, Mitsuuchi Y, Toda K, Miyahara K, Yokoyama Y, Nakao K, et al. Cloning of cDNA and genomic DNA for human cyto-chrome P-45011beta. FEBS Lett 1990;269:345–9.
12. Mulatero P, Schiavone D, Fallo F, Rabbia F, Veglio F. CYP11B2 gene polymorphisms in idiopathic hyperaldosteronism. Hypertension 2000;35:694–8.
13. Zhu H, Sagnella GA, Dong Y, Miller MA, Onipinla A, Makandu ND, et al. Contrasting associations between aldosterone synthase gene polymorphisms and essential hypertension in blacks and in whites. J Hypertens 2003;21:87–95.
14. White PC, Slutsker L. Haplotype analysis of CYP11B2. Endocr Res 1995;21:437–42.
15. Zucker IH. Novel mechanisms of sympathetic regulation in chronic heart failure. Hypertension 2006;48:1005–11.
16. Yu Y, Wei SG, Zhang ZH, Gomez-Sanchez E, Weiss RM, Felder RB. Does aldosterone upregulate the brain renin-angiotensin system in rats with heart failure? Hypertension 2008;51:727–33.
17. Lindley TE, Doobay MF, Sharma RV, Davisson RL. Superoxide is involved in the central nervous system activation and sympathoexcitation of myocardial infarction-induced heart failure. Circ Res 2004;94:402–9.
18. Huang BS, Ahmad M, Tan J, Leenen FHH. Sympathetic hyperactivity and cardiac dysfunction post-MI: different impact of specific CNS versus general AT1 receptor blockade. J Mol Cell Cardiol 2007;43:479–86.
19. Lal A, Veinot JP, Leenen FHH. Critical role of CNS effects of aldosterone in cardiac remodeling post-myocardial infarction in rats. Cardiovasc Res 2004;64:437–76.
20. Huang BS, Leenen FHH. Blockade of brain mineralocorticoid receptors or Na+ channels prevents sympathetic hyperactivity and improves cardiac function in rats post-MI. Am J Physiol 2005;288:H2491–7.
21. Huang SB, White AR, Ahmad M, Tan J, Jeng YA, Leenen HHF. Central infusion of aldosterone synthase inhibitor attenuates left ventricular dysfunction and remodeling in rats after myocardial infarction. Cardiovasc Res 2009;81:574–81.
22. MacFadyen RJ, Lee AF, Morton JJ, Pringle SD, Struthers AD. How often are angiotensin II and aldosterone concentrations raised during chronic ACE inhibitor treatment in cardiac failure? Heart 1999;82:57–61.
23. Cicoira M, Zanolla L, Rossi A, Golia G, Franceschini L, Cabrini G, et al. Failure of aldosterone suppression despite angiotensin-converting enzyme (ACE) inhibitor administration in chronic heart failure is associated with ACE DD genotype. J Am Coll Cardiol 2001;37:1808–12.
24. Schjoedt KJ, Andersen S, Rossing P, Tarnow L, Parving HH. Aldosterone escape during blockade of the renin–angiotensin–aldosterone system in diabetic nephropathy is associated with enhanced decline in glomerular filtration rate. Diabetologia 2004;47:1936–9.
25. Naruse M, Tanabe A, Sato A, Takagi S, Tsuchiya K, Imaki T, et al. Aldosterone breakthrough during angiotensin II receptor antagonist therapy in strokeprone spontaneously hypertensive rats. Hypertension 2002;40:28–33.
26. Pitt B, Zannad F, Remme WJ, Cody R, Castaigne A, Perez A, et al. The effect of spironolactone on morbidity and mortality in patients with severe heart failure. Randomized Aldactone Evaluation Study Investigators. N Engl J Med 1999;341:709–17.
27. Mantero F, Lucarelli G. Aldosterone antagonists in hypertension and heart failure. Ann Endocrinol (Paris) 2000;61:52–60.
28. Soberman JE, Weber KT. Spironolactone in congestive heart failure. Curr Hypertens Rep 2000;2:451–6.
29. Delyani JA. Mineralocorticoid receptor antagonists: the evolution of utility and pharmacology. Kidney Int 2000;57:1408–11.
30. Rousseau MF, Gurne O, Duprez D, Van Mieghem W, Robert A, Ahn S, et al. Beneficial neurohormonal profile of spironolactone in severe congestive heart failure: results from the RALES neurohormonal substudy. J Am Coll Cardiol 2002;40:1596–601.
31. Chai W, Garrelds IM, de Vries R, Batenburg WW, van Kats JP, Danser AH. Nongenomic effects of aldosterone in the human heart: interaction with angiotensin II. Hypertension 2005;46:701–6.
32. Fujita M, Minamino T, Asanuma H, Sanada S, Hirata A, Wakeno M, et al. Aldosterone nongenomically worsens ischemia via protein kinase C-dependent pathways in hypoperfused canine hearts. Hypertension 2005; 46:113–17.
33. Varo N, Etayo JC, Zalba G, Beaumont J, Iraburu MJ, Montiel C, et al. Losartan inhibits the post-transcriptional synthesis of collagen type I and reverses left ventricular fibrosis in spontaneously hypertensive rats. J Hypertens 1999;17:107–14.
34. Thai HM, Van HT, Gaballa MA, Goldman S, Raya TE. Effects of AT1 receptor blockade after a myocardial infarct on myocardial fibrosis, stiffness and contractility. Am J Physiol 1999;276:H873–80.
35. Pilon C, Mulatero P, Barzon L, Veglio F, Garonne C, Boscaro M, et al. Mutations in *CYP11B1* gene converting 11beta-hydroxylase into an aldosterone-producing enzyme are not present in aldosterone-producing adenomas. J Clin Endocrinol Metab 1999;84:4228–31.
36. Ulmschneider S, Müller-Vieira U, Mitrenga M, Hartmann RW, Oberwinkler-Marchais S, Klein CD, et al. Synthesis and evaluation of imidazolylmethylenetetrahydronaphthalenes and imidazolylmethyleneindanes: potent inhibitors of aldosterone synthase. J Med Chem 2005;48:1796–805.
37. Ulmschneider S, Müller-Vieira U, Klein CD, Antes I, Lengauer T, Hartmann RW. Synthesis and evaluation of (pyridylmethylene-) tetrahydronaphthalenes/-indanes and structurally modified derivatives: potent and selective inhibitors of aldosterone synthase. J Med Chem 2005;48:1563–75.
38. Heim R, Lucas S, Grombein MC, Ries C, Schewe EK, Negri M, et al. Overcoming undesirable CYP1A2 inhibition of pyridylnaphthalene-type aldosterone synthase inhibitors: influence of heteroaryl derivatization on potency and selectivity. J Med Chem 2008;51:5064–74.
39. Lucas S, Heim R, Negri M, Antes I, Ries C, Schewe EK, et al. Novel aldosterone synthase inhibitors with extended carbocyclic skeleton by a combined ligand-based and structure-based drug design approach. J Med Chem 2008;51:6138–49.
40. Voets M, Antes I, Scherer C, Müller-Vieira U, Biemel K, Barassin C, et al. Heteroaryl-substituted naphthalenes and structurally modified derivatives: selective inhibitors of CYP11B2 for the treatment of congestive heart failure and myocardial fibrosis. J Med Chem 2005;48:6632–42.
41. Voets M, Antes I, Scherer C, Müller-Vieira U, Biemel K, Marchais-Oberwinkler S, et al. Synthesis and evaluation of heteroaryl-substituted dihydronaphthalenes and indenes: potent and selective inhibitors of aldosterone synthase (CYP11B2) for the treatment of congestive heart failure and myocardial fibrosis. J Med Chem 2006;49:2222–31.
42. Lucas S, Heim R, Ries C, Schewe EK, Birk B, Hartmann WR. In vivo active aldosterone synthase inhibitors with improved selectivity: lead optimization providing a series of pyridine substituted 3,4-dihydro-1H-quinolin-2-one derivatives. J Med Chem 2008;51:8077–87.
43. Pinto-Bazurco Mendieta MA, Negri M, Jagusch C, Müller-Vieira U, Lauterbach T, Hartmann WR. Synthesis, biological evaluation, and molecular modeling of abiraterone analogues: novel CYP17 inhibitors for the treatment of prostate cancer. J Med Chem 2008; 1:5009–18.
44. Cerius2 version 4.8. San Diego, CA: Accelrys, Inc. (available at: http://www.accelrys.com/cerius2).
45. Leonard JT, Roy K. On selection of training and test sets for the development of predictive QSAR models. QSAR Comb Sci 2006;25:235–51.
46. Hopfinger AJ, Tokarsi JS. Three-dimensional quantitative structure-activity relationship analysis. In: Charifson PS, ed. Practical Applications of Computer-Aided Drug Design. New York: Marcel Dekker, 1997:105–64.
47. Fan Y, Shi LM, Kohn KW, Pommier Y, Weinstein JN. Quantitative structure-antitumor activity relationships of camptothecinanalogues: cluster analysis and genetic algorithm-based studies. J Med Chem 2001;44:3254–63.

48. Rogers D, Hopfinger AJ. Application of genetic function approximation to quantitative structure-activity relationship and quantitative structure-property relationship. J Chem Inf Comput Sci 1994;34:854–66.
49. Dunn WJ III, Rogers D. Genetic partial least squares in QSAR. In: Devillers J, ed. Genetic Algorithms in Molecular Modeling. London: Academic Press, 1996:109–30.
50. Hasegawa K, Miyashita Y, Funatsu K. GA strategy for variable selection in QSAR studies: GA-based PLS analysis of calcium channel antagonists. J Chem Inf Comput Sci 1997;37:306–10.
51. Snedecor GW, Cochran WG. Statistical Methods. New Delhi: Oxford & IBH Publishing Co. Pvt. Ltd., 1967.
52. Wold S, Eriksson L. Statistical validation of QSAR results. Validation tools. In: van de Waterbeemd H, ed. Chemometric Methods in Molecular Design. Weinheim: VCH, 1995:312–17.
53. Debnath AK. Quantitative structure-activity relationship (QSAR): A versatile tool in drug design. In: Ghose AK, Viswanadhan VN, eds. Combinatorial Library Design and Evaluation. New York: Marcel Dekker, 2001:73–129.
54. Roy K. On some aspects of validation of predictive QSAR models. Expert Opin Drug Discov 2007;2:1567–77.
55. Roy PP, Roy K. On some aspects of variable selection for partial least squares regression models. QSAR Comb Sci 2008;27:302–13.
56. Roy K, Roy PP. Comparative QSAR studies of CYP1A2 inhibitor flavonoids using 2D and 3D descriptors. Chem Biol Drug Des 2008;72:370–82.
57. Roy PP, Leonard JT, Roy K. Exploring the impact of the size of training sets for the development of predictive QSAR models. Chemom Intell Lab Syst 2008;90:31–42.
58. Eriksson L, Jaworska J, Worth AP, Cronin MTD, McDowell RM, Gramatica P. Methods for reliability and uncertainty assessment and for applicability evaluations of classification- and regression-based QSARs. Environ Health Perspect 2003;111:1361–75.