Available online at www.sciencedirect.com
ScienceDirect
Bioorganic & Medicinal Chemistry Letters 17 (2007) 2474–2482
Bioorganic & Medicinal Chemistry Letters

Prediction of binding affinities to β_1 isoform of human thyroid hormone receptor by genetic algorithm and projection pursuit regression

Yueying Ren,[a] Huanxiang Liu,[a] Shuyan Li,[a] Xiaojun Yao[a,b,*] and Mancang Liu[a]

[a]*Department of Chemistry, Lanzhou University, Lanzhou 730000, China*
[b]*State Key Laboratory of Applied Organic Chemistry, Lanzhou University, Lanzhou 730000, China*

Received 2 November 2006; revised 21 January 2007; accepted 9 February 2007
Available online 13 February 2007

Abstract—Quantitative structure–activity relationship (QSAR) has been applied to a set of thyroid hormone receptor β_1 (TRβ_1) antagonists, which are of special interest because of their potential role in safe therapies for nonthyroid disorders while avoiding the cardiac side effects. Using the calculated structural descriptors by CODESSA program, principal component analysis (PCA) was performed on the whole compounds to assist the separation of the data into the training set and the test set in QSAR analysis. Six molecular descriptors selected by genetic algorithm (GA) were used as inputs for a projection pursuit regression (PPR) study to develop a more accurate QSAR model. The PPR model performs well both in the fitting and prediction capacity. For the test set, it gave a predictive correlation coefficient (R) of 0.9450, root mean square error (RMSE) of 0.4498, and absolute average relative deviation (AARD) of 4.19%, respectively, confirming the ability of PPR for the prediction of the binding affinities of compounds to β_1 isoform of human thyroid hormone receptor (TRβ_1).

Thyroid hormones can exert diverse effects on growth, development, and homeostasis in mammals.[1] There are two subtypes of the thyroid hormone receptors, TRα and TRβ. TRα_1 is suggested to mediate most effects on the heart, whilst the most actions of the hormones on the liver and other tissues are mediated more through the β-forms of the receptor.[2,3] Based on the recently described novel concept of 'passive antagonism' of estrogen receptors,[4] a series of TR ligands selective for TRβ_1 have been synthesized.[5–8] These thyroid hormone receptor antagonists could lead to safe therapies for nonthyroid disorders while avoiding the cardiac side effects and, therefore, could be used as a short term supplemental therapy to the conventional treatments.[7] Also, corresponding structure–activity relationships (SARs) were studied to identify compounds with enhanced TRβ_1 affinity and selectivity.[5–8]

Quantitative structure–activity relationship (QSAR) study is a major field of research in medicinal chemistry and drug design. At present, various QSAR studies have been successfully carried out in this field.[9–13] For drugs binding to TRβ_1, however, only Vedani et al. performed a QSAR study which showed a certain predictive capability and was proved to be suitable for the in silico identification of adverse effects triggered by drugs and chemicals.[13]

The present paper attempts to develop QSAR model of TRβ_1 receptor binding using the combination of genetic algorithm and projection pursuit regression (PPR) method, developed by Friedman.[14] The goal is to develop a reliable QSAR model to correlate the TRβ_1 binding affinities of compounds with structural information, which will be very useful for the prediction of the binding affinity to TRβ_1 and for the design of new compounds with good binding affinity.

Projection pursuit (PP) is a powerful tool for seeking interesting projections of high-dimensional data into lower-dimensional space and, therefore, can overcome the curse of dimensionality. At present, it has been applied successfully to tackle some chemical problems.[15–17] The basic theory of PP can be found in details

Keywords: QSAR; β_1 Isoform of human thyroid hormone receptor; Principle component analysis; Genetic algorithm; Projection pursuit regression.
* Corresponding author. Tel.: +86 931 891 2578; fax: +86 931 891 2582; e-mail: xjyao@lzu.edu.cn

0960-894X/$ - see front matter © 2007 Elsevier Ltd. All rights reserved.
doi:10.1016/j.bmcl.2007.02.025

in Ref. 14. Therefore, we only give a brief description here. Given a data set $(\mathbf{X}_1, \ldots, \mathbf{X}_n)$, $\mathbf{X} \in \mathrm{IR}^{\mathbf{k}}$ are **k**-dimensional matrix $(\mathbf{k} \times \mathbf{n})$, where **k** is the number of observed variables and **n** is the number of units, and an **m**-dimensional orthonormal matrix $\mathbf{A}(\mathbf{m} \times \mathbf{k})$, the $(\mathbf{m} \times \mathbf{n})$ matrix $\mathbf{Z} = \mathbf{AX}$ represents the coordinates of the projection data onto the **m**-dimensional $(\mathbf{m} < \mathbf{k})$ space spanned by the rows of **A**. As such projections are infinite, it is important to have a technique to pursue a finite sequence of projections that can reveal the most interesting structures of the data. Projection pursuit (PP) is such powerful tool that combines both ideas of projection and pursuit.[14] In a typical regression problem, projection pursuit regression (PPR) aims to approximate the regression function $f(x)$ by a finite sum of ridge functions with suitable choices of a_i and g_i:

$$g^{(p)}(x) = \sum_{i=1}^{p} g_i(\alpha_i^{\mathrm{T}} x), \tag{1}$$

where a_i are $\mathbf{m} \times \mathbf{n}$ orthonormal matrices, p is the number of ridge functions.

In this study, we have collected a data set of 80 molecules whose affinity for TRβ_1 was reported.[5–8] The affinities were represented using $\mathbf{pIC}_{50}$, which ranges between 4 and 11 with an average value of 8.0592 log unit; shown in Table 1 together with the molecular structures. The structures of these molecules were drawn with **ISIS Draw 2.3** program.[18] The molecular descriptors were calculated using CODESSA software.[19] A full list of the calculated molecular descriptors as well as types is shown in Supplementary Material 1. Principal component analysis (PCA) with these calculated molecular descriptors was performed on the whole data with two aims: first, to detect the homogeneities in the data set, that is, to identify possible outliers and clusters and, second, to show spatial location of compounds and to assist the separation of the data set into representative training set and the test set.

Here, PCA gives three significant PCs (eigenvalues > 1), which explains 64.99% of the variation in the data (40.42%, 14.50%, and 10.07%, respectively). Figure 1 shows the distribution of compounds over the three first principal components. It can be seen clearly that the compounds were clustered into two main clusters (compounds **14**–**24** and the rest). This was due to the structure deviation (shown in Table 1) among these compounds. In addition, compound **11** was found to be farther away from the majority of compounds.

Based on the results of PCA, the whole data were split into two subsets: the training set containing 67 compounds to develop the model and the test set of the remained 13 compounds to evaluate the model performance. As can be seen from Figure 1, the compounds in each subset seem to be relatively well balanced over the space of the principal components. It can confirm the representative ability of the compounds in each subset during data set splitting.

In QSAR studies, it is believed that bioactivities were functions of those related descriptors. Irrelevant descriptors, however, can cause the system to focus attention on the idiosyncrasies of the individual samples and lose sight of the broad picture that is essential for generalization beyond the training set. For this reason, variable selection is very important. Genetic algorithm (GA), as a powerful tool in searching the most suitable parameters for a regression model,[20] was applied in this study. Genetic algorithm is a stochastic optimization method developed by Holland et al. in the 1960s.[21] It is inspired and named by biological process of inheritance, mutation, natural selection, and the genetic crossover that occurs when parent mate to produce offspring.[22] Here, it is used to relate binding affinity with different combinations of descriptors with appropriate physicochemical significance in order to find out the best multiple regression. R^2, the explained variance, was used as the fitness function calculated below:

$$R^2 = 1 - \frac{\sum_{i=1}^{n} (y_{ke} - y_{kp})^2}{\sum_{i=1}^{n} (\bar{y} - y_{ke})^2} \tag{2}$$

Where k represents kth molecule, y_{ke} is the desired output (experimental property), y_{kp} is the actual output of the models, $\bar{y}$ is the average values of the studied property, and n is the number of compounds in analyzed set. When adding another descriptor did not improve the fitness significantly, it was determined that the optimal set was obtained.

In this work, six descriptors were selected using GA whose values are shown in Supplementary Material 2. Using these variables, a linear regression model was built and is shown in Table 2. The statistical parameters were: $n = 67$, $R^2 = 0.7721$, adjusted $R^2 = 0.7498$, cross-validated $R^2 = 0.7320$, $s = 0.7009$, $F = 33.8778$ (95% confidence level), and $p < 0.0001$. The interpretation of the physical significance of each parameter in the model is not straightforward as there are both linear (original calculated structural descriptors by CODESSA program) and nonlinear (in this case, the squared terms of the original descriptors) variables. $C_{\mathrm{min(H)}}$ means the minimum net atomic charge for a H atom. It depends directly on the quantum-chemically calculated charge distribution in the molecules, and therefore describes the polar interactions between molecules or their chemical nondirectional reactivity index.[23] $E_{\mathrm{Mainval(H)}}$ is the minimum atomic valence state energy for a H atom and can be used as a measure of nonspecific interactions between the antagonists and the TRβ_1 receptor. $E_{\mathrm{min,\,tot(C–O)}}$ means the minimum total interaction energy for a C–O bond and is calculated as the summation of two terms, the minimum electronic exchange energy for a C–O bond and the minimum coulomb interaction energy for a C–O bond; it may be related to the conformational (rotational and inversional) changes or atomic reactivity in the molecule.[24] $(\mathrm{RMW})^2$ is the square of relative molecular weight; this parameter is related to molecular size. As this value increases, the size and the hydrophobicity of a molecule increase, and then the affinity of the molecule to TRβ_1 receptor increases. $(\mathrm{BO}_{\sigma\pi})^2$ and $(\mathrm{BO}_{\mathrm{Avg(C)}})^2$, which are

Table 1. List of structures, experimental and predicted binding affinities as well as the corresponding errors of the studied compounds

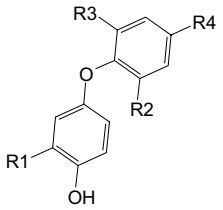
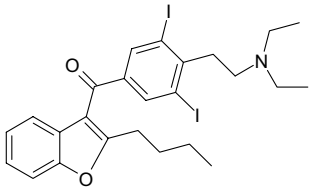
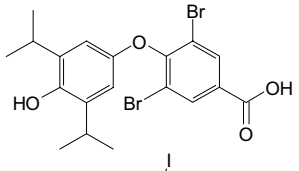
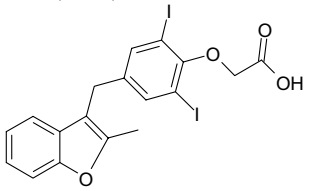
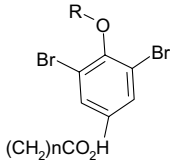
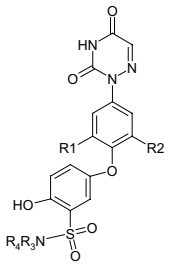
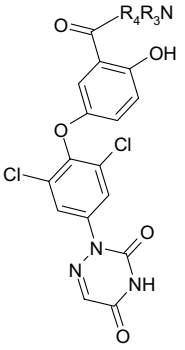
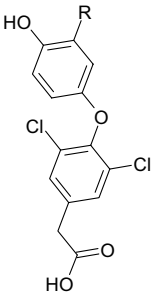
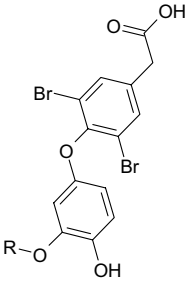
Structure	Compound	Substituents	pIC_{50}				
			Exp.	Linear	Error	PPR	Error
		**R^4**					
	1*	$-CH_2CH(NH_2)CO_2H$	9.585	10.0249	−0.4399	9.8431	−0.2581
	2	$-CH_2CO_2H$	10.3188	10.3995	−0.0807	10.711	−0.3922
	3	$-CH_2CH(NH_2)CO_2H$	9.9586	9.991	−0.0324	10.2668	−0.3082
	4	$-(CH_2)_2CO_2H$	10.7213	10.2139	0.5074	10.2013	0.52
	5	$-(CH_2)_2CO_2H$	10.6021	8.925	1.6771	9.9112	0.6909
1,**2**,**4**: $R^1 = R^2 = R^3 = I$;	**6**	$-CO_2H$	8.6778	7.6919	0.9859	8.1669	0.5109
3,**5**–**10**: R^1 = Isopropyl;	**7***	$-CH_2CO_2H$	10.0223	9.0959	0.9264	9.9204	0.1019
3: $R^2 = R^3 = I$;	**8**	$-CO_2H$	7.6778	7.2734	0.4044	7.4688	0.209
5–**7**: $R^2 = R^3 = Br$;	**9**	$-CH_2CO_2H$	8.9586	9.3311	−0.3725	9.3108	−0.3522
8–**10**: $R^2 = R^3 = Cl$	**10**	$-(CH_2)_2CO_2H$	9.8239	9.2198	0.6041	9.4551	0.3688
	**11**		6.2218	6.0995	0.1223	6.3478	−0.126
	**12**		6.041	7.0917	−1.0507	6.7977	−0.7567
	**13**		5.2924	6.0824	−0.79	5.6174	−0.325
		R					
	**14**	Et	4.4949	4.8366	−0.3417	4.5705	−0.0756
	15	$-CH_2(CH_2)_2CH_3$	5.7212	5.3744	0.3468	5.4263	0.2949
	16	$-CH_2(CH_2)_3CH_3$	5.2441	5.7471	−0.503	5.2387	0.0054
	17*	$-CH_2CH_2CH(CH_3)_2$	5.7447	6.6248	−0.8801	5.8238	−0.0791
	18	$-CH_2(CH_2)_4CH_3$	5.8239	5.6481	0.1758	5.7095	0.1144
	19	$-CH_2CH(CH_2CH_3)_2$	6.0969	7.0741	−0.9772	6.2584	−0.1615
	20	$-CH_2$-Cyclohexyl	6.1135	6.3365	−0.223	6.5531	−0.4396
14–**20**: $n = 1$;	**21***	$-CH_2CH_2CH(CH_3)_2$	5.7447	6.5601	−0.8154	5.6115	0.1332
21–**24**: $n = 2$	**22**	$-CH_2(CH_2)_4CH_3$	6.4437	5.8196	0.6241	6.469	−0.0253
	23	$-CH_2CH(CH_2CH_3)_2$	6.0915	6.4799	−0.3884	5.8582	0.2333
	24	$-CH_2$–Cyclohexyl	6.7212	6.2749	0.4463	6.5997	0.1215
		**$N–R^3R^4$**					
	25	Piperidinyl	8.8916	9.0902	−0.1986	8.5051	0.3865

Table 1 (*continued*)

Structure	Compound	Substituents	pIC_{50}				
			Exp.	Linear	Error	PPR	Error
	26	Piperidinyl	9.4405	9.0785	0.362	9.1536	0.2869
	27	Piperidinyl	10.2534	10.2791	−0.0257	10.1341	0.1193
	28	4-Methylpiperidinyl	7.4405	9.0415	−1.601	8.5171	−1.0766
	29	Cyclohexylamino	8.8916	9.321	−0.4294	9.1374	−0.2458
	30	Cyclobutylamino	8.9523	9.0278	−0.0755	8.7755	0.1768
	31	Anilinyl	9.5544	8.4191	1.1353	9.6988	−0.1444
25: $R_1 = R_2$= Me;	**32**	Indolinyl	9.1927	8.8167	0.376	8.9411	0.2516
26, **28**: R_1 = Me, R_2 = Cl;	**33**		9.063	8.4581	0.6049	9.3983	−0.3353
27, **29**–**34**: $R_1 = R_2$= Cl	**34***		9.1072	8.052	1.0552	8.5813	0.5259
		N–$\mathbf{R^3R^4}$					
	**35**	Piperidinyl	8.2756	9.1722	−0.8966	8.1365	0.1391
	36	Piperidinyl	8.7551	8.8767	−0.1216	9.2631	−0.508
	37	Cyclohexylamino	9.7093	8.7455	0.9638	9.763	−0.0537
	38	Cyclobutylamino	8.7482	9.1859	−0.4377	8.8337	−0.0855
	39*	Morpholinyl	8.6971	8.0068	0.6903	8.5418	0.1553
	40	(*R*)-(+)-Bornyl	8.9862	9.0263	−0.0401	8.9954	−0.0092
	41*		9.7762	9.5327	0.2435	9.3458	0.4304
	42		9.9523	9.2187	0.7336	9.4352	0.5171

(*continued on next page*)

Table 1 (*continued*)

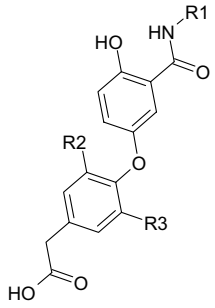
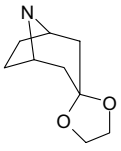
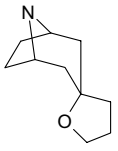
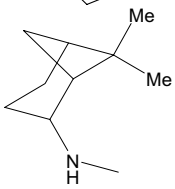
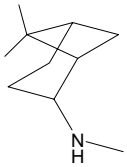
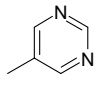
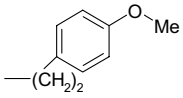
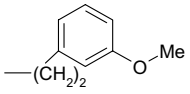
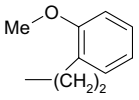
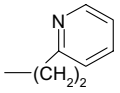
Structure	Compound	Substituents	pIC_{50}				
			Exp.	Linear	Error	PPR	Error
		X					
	**43**	H2	8.5376	7.6676	0.87	7.7134	0.8242
	44	$-CF_3$	8.4685	8.2631	0.2054	8.0971	0.3714
	45	3-CF_3	9.0269	8.4452	0.5817	8.7181	0.3088
	46	4-CF_3	7.1739	8.1946	−1.0207	7.7157	−0.5418
	47	3-Et	9.699	8.6557	1.0433	8.9556	0.7434
	48	3-Isopropyl	8.3979	9.0361	−0.6382	8.8447	−0.4468
	49	3-Ph	7.6778	7.9866	−0.3088	8.0761	−0.3983
	50	3-OMe	7.9586	7.5655	0.3931	7.5655	0.3931
	51	3-$OCHF_2$	9.1135	7.2591	1.8544	8.8732	0.2403
	52	3-OCF_3	8.4685	8.4243	0.0442	8.486	−0.0175
	53*	2-OH	7.301	7.0286	0.2724	6.9986	0.3024
	54	3-OH	7.1249	7.727	−0.6021	7.6009	−0.476
	55	4-OH	7.0605	7.7098	−0.6493	7.4717	−0.4112
	**56**		6.857	7.1876	−0.3306	7.0444	−0.1874
	57		7.3872	7.1705	0.2167	6.8832	0.504
	58*		6.9066	7.2561	−0.3495	7.2226	−0.316
	59	R= 	7.1739	6.8122	0.3617	7.2108	−0.0369
	60	R= 	7.5528	7.0907	0.4621	7.3536	0.1992
	61	*n*-C_6H_{11}	8.1135	8.5727	−0.4592	8.6075	−0.494
		X					
R= 	**62**	H	8.224	8.0826	0.1414	7.2809	0.9431
	63	2-CF_3	8.1175	8.927	−0.8095	8.4985	−0.381
	64	3-CF_3	8.9586	9.2042	−0.2456	9.1399	−0.1813
	65*	4-CF_3	8.0223	9.3069	−1.2846	9.1533	−1.131
	66	3-Ph	7.2218	8.1665	−0.9447	7.9062	−0.6844

Table 1 (*continued*)

Structure	Compound	Substituents	pIC_{50}				
			Exp.	Linear	Error	PPR	Error
R= 	**67**	2-OH	7.0809	7.6493	−0.5684	7.6453	−0.5644
	68	3-OH	7.3279	7.441	−0.1131	7.1658	0.1621
	69	4-OH	8.2218	7.7021	0.5197	7.9031	0.3187
	No.	**R^1**					
R= 	**70***	Ph	7.4559	7.9231	−0.4672	7.6579	−0.202
	71	Bz	8.3768	8.6483	−0.2715	8.2261	0.1507
	72	$(CH_2)_2Ph$	9.2076	8.835	0.3726	8.9582	0.2494
	73	$(CH_2)_3Ph$	8.1675	8.5882	−0.4207	8.4141	−0.2466
	74	$(CH_2)_4Ph$	7.7447	8.6959	−0.9512	8.5772	−0.8325
	75		7.4318	7.7392	−0.3074	7.6646	−0.2328
	76*		8.3188	7.705	0.6138	7.5927	0.7261
	77		8.0706	7.7392	0.3314	7.6646	0.406
70–79: $R^2 = R^3 = Br$	**78**	CH_2CHPh_2	9.3279	8.7442	0.5837	8.6721	0.6558
80: $R^2 = R^3 = Cl$	**79**		7.1739	7.9617	−0.7878	7.0367	0.1372
	80*	$(CH_2)_2Ph$	8.4559	8.1335	0.3224	8.3812	0.0747

* Test set.

the squares of maximum SIGMA-PI bond order ($BO_{\sigma\pi}$) and the average bond order of a C atom $BO_{Avg(C)}$, respectively, are related to the strength of intramolecular bonding interactions and characterize the stability of the

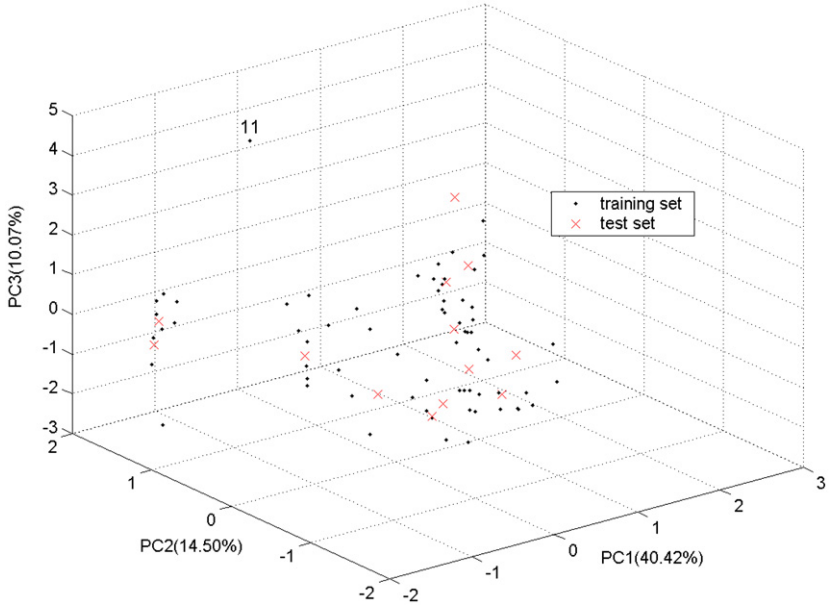


Figure 1. Principal component analysis of the structural descriptors for the data set.

Table 2. Descriptors, abbreviation, their coefficients, t test values in the linear regression model, and their variation inflation factors

Abbreviations	Unstandardized coefficients		Standardized coefficients	VIF
	B	Error	Beta	
Constant	53.028	8.204		
$C_{\min(H)}$	−17.952	3.717	−0.341	1.311
$E_{\text{Mainval(H)}}$	−7.939	1.049	−0.521	1.248
$E_{\min,\text{tot(C–O)}}$	1.712	0.166	0.669	1.105
$(\text{RMW})^2$	0.009	0.002	0.347	1.183
$(\text{BO}_{\sigma\pi})^2$	411.779	105.818	0.285	1.408
$(\text{BO}_{\text{Avg(C)}})^2$	−15.169	2.124	−0.534	1.475

molecules, their conformational flexibility, and other valency-related properties.[25] The influence of individual descriptors on the binding affinity is different from each other and can be reflected by the standardized regression coefficients. As can be seen from Table 1, three descriptors, that is, $E_{\min,\text{tot(C–O)}}$, $(\text{BO}_{\text{Avg(C)}})^2$, and $E_{\text{Mainval(H)}}$, play major role in the binding process. The negative coefficient of any descriptor in the regression function means increasing this descriptor will decrease the pIC_{50} values and therefore increase the binding ability. On the contrary, an increase of the descriptor with positive coefficient would lead to a decrease of the binding ability to $\text{TR}\beta_1$ receptor.

Mechanistic interpretations of individual terms in a regression equation may not always result in an entirely meaningful conclusion; however, the descriptors as a group gave the best explanation of the variation of the dependent variable. In summary, the factors influencing the binding strengths of antagonists to $\text{TR}\beta_1$ receptor are mainly the electronic, steric, and, to a lesser extent, hydrophobic of the molecules.

In addition, the intercorrelations among these variables were checked with their variation inflation factor (VIF) values (calculated as $1/(1-r^2)$, where r is the correlation coefficient of multiple regression between one variable and the others in the model). If VIF equals to 1.0, no intercorrelation exists for each variable; if VIF falls into the range of 1.0–5.0, the related model is acceptable; and if VIF is larger than 10.0, the related model is unstable and recheck is necessary.[26] In this equation, all VIF values are far less than 5 (shown in Table 2) indicating that the obtained model has obvious statistical significance.[26]

Using this model, the binding affinities were predicted (shown in Table 1) but the results were not satisfactory. For the test set, it gave R^2 as 0.7233 indicating a poor predictive ability to these compounds.

For that reason, the PPR method was used to perform an effective project of the six descriptors from high-dimensional to the lower-dimensional space and, to perform a regression in this lower-dimensional space correlating the $\text{TR}\beta_1$ binding affinities of compounds with structural information. PPR algorithm was performed using R script. In this investigation, the PPR algorithm proposed by Friedman was used where g_i are found by smoothing operation that entails a backfitting. Three parameters 'nterms', 'optlevel', and 'span' need to be determined. The parameter 'nterms' controls the number of variables to be entered in the model, 'optlevel' means the levels of optimization which differ in how thoroughly the models are refitted during this process, and 'span' defines the fraction of the observations in the span of the running lines smoother. Here, optimal 'nterms', 'optlevel', and 'span' were determined as 6, 3, and 0.17, respectively.

In order to check the performance of the obtained model, we compared the results of PPR with those of the above linear model (shown in Table 3). R^2 was used to evaluate the fitness ability, whereas root mean square error (RMSE) and the absolute average relative deviation (AARD) were used to evaluate the model's predictive performance; calculated as follows:

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{n}(y_{ke}-y_{kp})^2}{n}}, \tag{3}$$

$$\text{AARD} = \frac{100}{n}\sum_{i=1}^{n}\frac{|y_{ke}-y_{kp}|}{y_{ke}}. \tag{4}$$

The smaller the RMSE and AARD, the more accurate the proposed model is. From Table 2, it can be seen that PPR model more stable and more predictive than the linear model.

We further analyzed the detailed prediction results by PPR (shown in Table 1 and Fig. 2). As can be seen from this figure, the predicted values are in well agreement with the experimental values for most compounds except that **12**, **28**, **43**, **62**, **66**, and **74** in the training set and **65** and **76** in the test set showed larger relative deviation as 12.53%, 14.47%, −9.65%, −11.47%, 9.47%, 10.75%, 16.59%, and −8.73%, respectively. Figure 3 shows proportion of compounds within a given deviation from the experimental pIC_{50} by PPR. It can be seen from Figure 3 that if one can tolerate 0.5 log unit of the absolute error, the model can correctly predict the affinity for 77.50% of the compounds for the whole data set with 77.61% and 76.92% for the training and test set, respectively. If one can tolerate an absolute error of 1 log unit, the model can correctly predict the affinity for 97.5% compounds with 98.51% and 92.31% for the training and test set, respectively. Only compound **28** in the training set and compound **65** in the test set showed absolute deviations more than one log unit as 1.08 and 1.13, respectively.

Table 3. The statistical parameters of the linear and PPR model obtained for the six descriptors selected by GA

Model	Data set	R^2	RMSE	AARD (%)
Linear	Training	0.7721	0.6633	6.87
	Test	0.7233	0.7172	8.28
PPR	Training	0.9097	0.4176	4.32
	Test	0.8928	0.4498	4.19

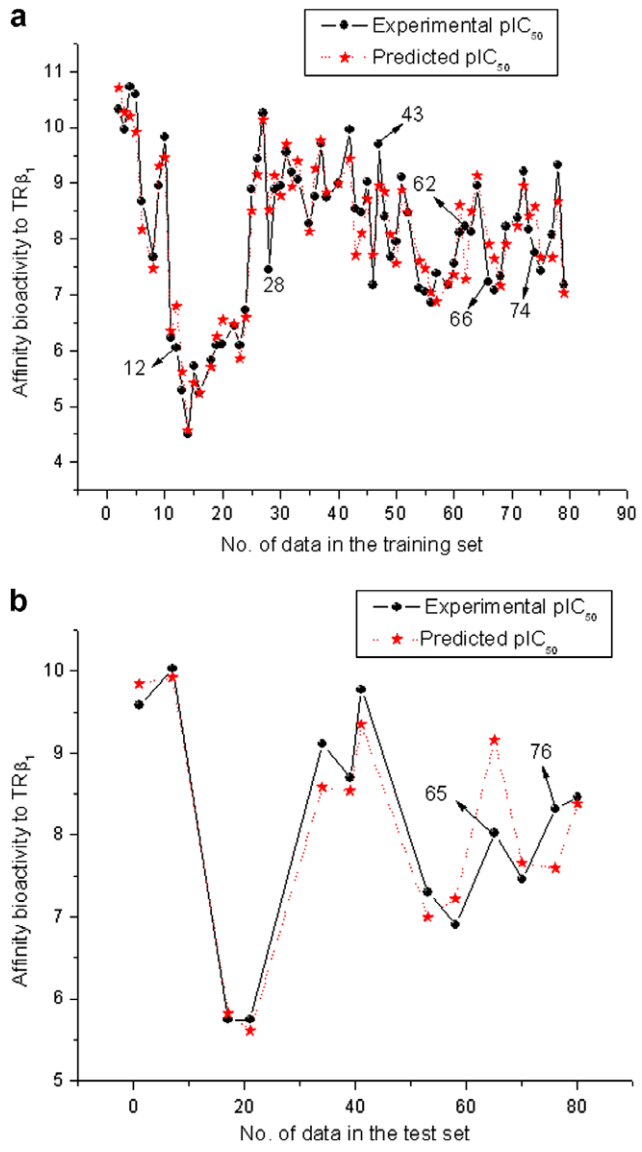


Figure 2. Predicted versus experimental pIC_{50} values for the data sets by PPR.

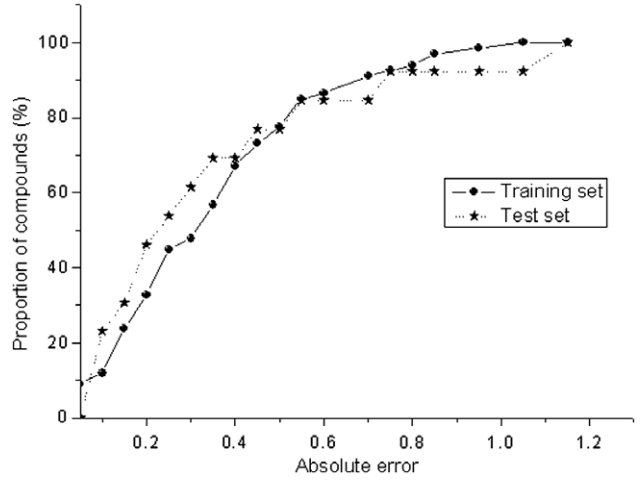


Figure 3. Proportion of compounds within a given deviation from the experimental pIC_{50} by PPR.

In summary, a projection pursuit regression (PPR) based QSAR study was developed in this work. The above calculation results, both correlations and predictions, show that our model is reliable with good predictive accuracy. The QSAR models developed in this study will be very useful for the prediction of the binding affinity of the newly synthesized compounds and also for the design of new compounds with good binding affinity.

Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.bmcl.2007.02.025.

References and notes

1. Lazar, M. A. *Endocr. Rev.* **1993**, *14*, 184.
2. Forrest, D.; Vennström, B. *Thyroid* **2000**, *10*, 41.
3. Takeda, K.; Sakurai, A.; DeGroot, L. J.; Refetoff, S. *J. Clin. Endocrin. Metab.* **1992**, *74*, 49.
4. Shiau, A. K.; Barstad, D.; Radek, J. T.; Meyers, M. J.; Nettles, K. W.; Katzenellenbogen, B. S.; Katzenellenbogen, J. A.; Agard, D. A.; Greene, G. L. *Nat. Struct. Biol.* **2002**, *9*, 359.
5. Ye, L.; Li, Y. L.; Mellström, K.; Mellin, C.; Bladh, L.-G.; Koehler, K.; Garg, N.; Collazo, A. M. G.; Litten, C.; Husman, B.; Persson, K.; Ljunggren, J.; Grover, G.; Sleph, P. G.; George, R.; Malm, J. *J. Med. Chem.* **2003**, *46*, 1580.
6. Hangeland, J. J.; Doweyko, A. M.; Dejneka, T.; Friends, T. J.; Devasthale, P.; Mellström, K.; Sandberg, J.; Grynfar, M.; Sack, J. S.; Einspahr, H.; Färnegårdh, M.; Husman, B.; Ljunggren, J.; Koehler, K.; Sheppard, C.; Malm, J.; Ryono, D. E. *Bioorg. Med. Chem. Lett.* **2004**, *14*, 3549.
7. Hedfors, Å.; Appelqvist, T.; Carlsson, B.; Bladh, L.-G.; Litten, C.; Agback, P.; Grynfarb, M.; Koehler, K. F.; Malm, J. *J. Med. Chem.* **2005**, *48*, 3114.
8. Dow, R. L.; Schneider, S. R.; Paight, E. S.; Hank, R. F.; Chiang, P.; Cornelius, P.; Lee, E.; Newsome, W. P.; Swick, A. G.; Spitzer, J.; Hargrove, D. M.; Patterson, T. A.; Pandit, J.; Chrunyk, B. A.; LeMotte, P. K.; Danley, D. E.; Rosner, M. H.; Ammirati, M. J.; Simons, S. P.; Schulte, G. K.; Tate, B. F.; DaSilva-Jardine, P. *Bioorg. Med. Chem. Lett.* **2003**, *13*, 379.
9. Manivannan, E.; Prasanna, S. *Bioorg. Med. Chem. Lett.* **2005**, *15*, 4496.
10. González-Dıaz, H.; Uriartea, E. *Bioorg. Med. Chem. Lett.* **2005**, *15*, 5088.
11. Guo, W. M.; Hu, X. F.; Chu, N. P.; Yin, C. S. *Bioorg. Med. Chem. Lett.* **2006**, *16*, 2855.
12. Deswal, S.; Roy, N. *Eur. J. Med. Chem.* **2006**, *41*, 552.
13. Vedani, A.; Dobler, M.; Lill, M. A. *Pharmacol. Toxicol.* **2006**, *99*, 187.
14. Friedman, J. H. *J. Am. Stat. Assoc.* **1987**, *82*, 249.
15. Kvalheim, O. V.; Liang, Y. Z. *Anal. Chem.* **1992**, *64*, 936.
16. Hu, Q. N.; Liang, Y. Z.; Peng, X. L.; Yin, H.; Fang, K. T. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 437.
17. Liu, H. X.; Yao, X. J.; Liu, M. C.; Hu, Z. D.; Fan, B. T. *Talanta* **2007**, *71*, 258.
18. ISIS Draw2.3 (1990–2000) MDL Information Systems, Inc.
19. Katritzky, A. R.; Lobanov, V. S.; Karelson, M., CODESSA Version 2.0 Reference Manual 1995–1997.
20. Rogers, D.; Hopfinger, A. J. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 854.

21. Holland, J. H. In *Adaptation in natural and artificial systems*; Arbor, A., Ed.; University of Michigan Press: USA, 1975.
22. Goldberg, D. E. *Genetic Algorithms in Search, Optimization and Machine Learning*; Addison-Wesley Publishing Company: Reading, MA, 1989, 412.
23. Kikuchi, O. *Quant. Struct.-Act. Relat.* **1993**, *12*, 246.
24. Strouf, O. *Chemical Pattern Recognition*; Wiley: New York, 1986.
25. Sannigrahi, A. B. *Adv. Quant. Chem.* **1992**, *23*, 301.
26. Famini, G. R.; Penski, C. A.; Wilson, L. Y. *J. Phy. Org. Chem.* **1992**, *5*, 395.