



Article

# Structural Investigation for Optimization of Anthranilic Acid Derivatives as Partial FXR Agonists by *in Silico* Approaches

Meimei Chen <sup>1,\*</sup>, Xuemei Yang <sup>1</sup>, Xinmei Lai <sup>1</sup>, Jie Kang <sup>1</sup>, Huijuan Gan <sup>1</sup> and Yuxing Gao <sup>2</sup>

<sup>1</sup> College of Traditional Chinese Medicine, Fujian University of Traditional Chinese Medicine, Fuzhou 350122, China; mei\_tcm@163.com (X.Y.); keerer1990@163.com (X.L.); meinuohua2007@163.com (J.K.); ghjzyz@163.com (H.G.)

<sup>2</sup> College of Chemistry and Chemical Engineering, Xiamen University, Xiamen 361005, China; gaoxingchem@xmu.edu.cn

\* Correspondence: chenmeimei1984@163.com; Tel.: +86-591-2286-1513

Academic Editors: Humberto González-Díaz, Roberto Todeschini and Alejandro Pazos Sierra

Received: 9 March 2016; Accepted: 5 April 2016; Published: 8 April 2016

**Abstract:** In this paper, a three level *in silico* approach was applied to investigate some important structural and physicochemical aspects of a series of anthranilic acid derivatives (AAD) newly identified as potent partial farnesoid X receptor (FXR) agonists. Initially, both two and three-dimensional quantitative structure activity relationship (2D- and 3D-QSAR) studies were performed based on such AAD by a stepwise technology combined with multiple linear regression and comparative molecular field analysis. The obtained 2D-QSAR model gave a high predictive ability ( $R^2_{\text{train}} = 0.935$ ,  $R^2_{\text{test}} = 0.902$ ,  $Q^2_{\text{LOO}} = 0.899$ ). It also uncovered that number of rotatable single bonds (b\_rotN), relative negative partial charges (RPC<sup>-</sup>), opra's lead-like (opr\_leadlike), subdivided van der Waal's surface area (SlogP\_VSA2) and accessible surface area (ASA) were important features in defining activity. Additionally, the derived 3D-QSAR model presented a higher predictive ability ( $R^2_{\text{train}} = 0.944$ ,  $R^2_{\text{test}} = 0.892$ ,  $Q^2_{\text{LOO}} = 0.802$ ). Meanwhile, the derived contour maps from the 3D-QSAR model revealed the significant structural features (steric and electronic effects) required for improving FXR agonist activity. Finally, nine newly designed AAD with higher predicted EC<sub>50</sub> values than the known template compound were docked into the FXR active site. The excellent molecular binding patterns of these molecules also suggested that they can be robust and potent partial FXR agonists in agreement with the QSAR results. Overall, these derived models may help to identify and design novel AAD with better FXR agonist activity.

**Keywords:** QSAR; docking; FXR; anthranilic acid derivatives

## 1. Introduction

Farnesoid X receptor (FXR) is a nuclear receptor expressed in liver, gall bladder, intestine, kidney, and adrenal glands. It regulates important physiological roles in various metabolic pathways involved in bile acid, triglyceride, and glucose homeostasis. Now, FXR has become an attractive target for treating a wide range of metabolic diseases, including diabetes, cholestasis, liver fibrosis, and inflammatory bowel diseases [1–3]. Therefore, a number of synthetic steroidal and nonsteroidal FXR agonists have been developed so far. 6-ethyl-chenodeoxycholic acid20 (6-ECDCA) and GW4064 were the most important and widely used steroidal and nonsteroidal FXR agonists [4]. They both constitute full FXR agonists with low nanomolar EC<sub>50</sub> values of 85 nM and 0.9 M in a reporter gene assay [5], respectively. In addition, a recent study indicated that GW4064 was active on several off-targets [6]. Considering that metabolic diseases require a stable long-term therapy, well tolerated and low toxicity FXR agonists are predominantly required that can be applied over long time. Moreover,

full activation of a ligand activated transcription factor may cause many side effects in long-term treatment [7]. Therefore, new potent partial FXR agonists aimed at providing a stable long-term therapy for metabolic diseases have attracted more and more attention nowadays.

Recently, a novel series of partial FXR agonists based on anthranilic acid skeleton have been reported by Merk *et al.* [8,9]. The continued interest in the development of more potent partial FXR agonists prompted us to explore the relationship between structures of AAD and FXR agonist activity. Here, quantitative structure activity relationship (QSAR) methods were introduced to guide lead optimization and study the action mechanism for partial FXR agonists. In this QSAR method, the bioactivity of compounds can be predicted by a mathematical model between physicochemical properties and bioactivity. The mathematical model can be achieved by many general algorithms such as linear and nonlinear algorithms or other new methods such as the spectral-structure activity relationship algorithm [10,11]. This QSAR method has become very useful and is widely applied in many fields for predicting compound properties [12,13], including physical property prediction, biological activity prediction, and toxicity prediction.

To the best of our knowledge, no QSAR study has yet been reported in AAD as FXR agonists so far. In this paper, we attempted to investigate the significant structural and physicochemical features required for improving biological activity and to obtain highly predictive 2D- and 3D-QSAR models so as to assist in the design of new potent partial FXR agonists. Firstly, a stepwise technology combined with multiple linear regression (MLR) was applied to develop predictive 2D-QSAR models for uncovering physicochemical features on FXR activity of AAD. Subsequently, a 3D-QSAR study was also performed to obtain more understanding with respect to chemical structures and biological activity using comparative molecular field analysis (CoMFA). The CoMFA model can provide identification of regions in space where the interactive fields may influence the biological activities in the form of contour maps, which would generate graphical visualization of crucial steric and electrostatic features in 3D Cartesian space [14]. Finally, some important observations were also made during the study concerning nine newly designed AAD with high predicted bioactivity and their interactions with the FXR active site by molecular docking. Molecular docking aims to predict the binding-conformation of ligands to the appropriate target binding site. The success of a docking program depends on two components: the search algorithm and the scoring function. A variety of conformational search strategies have been reported such as systematic or stochastic search or genetic algorithms or simplified molecular input line entry system conformation [15]. Most scoring functions are physics-based molecular mechanics force fields that estimate the energy of the pose within the binding site.

## 2. Results and Discussion

### 2.1. Two-Dimensional Quantitative Structure Activity Relationship (2D-QSAR) Models

#### 2.1.1. Multiple Linear Regression Modeling

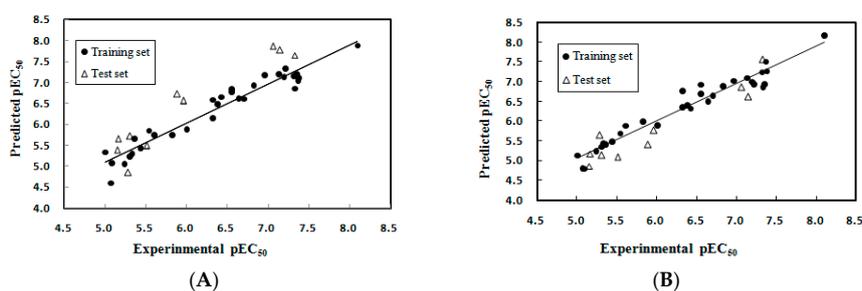
After stepwise multiple linear regression (SW-MLR) was performed, the best linear model was generated with five molecular descriptors. The obtained MLR model was given as follows:

$$\text{pEC}_{50} = (0.016 \pm 0.002)\text{ASA} + (14.001 \pm 4.041)\text{RPC}^- - (0.049 \pm 0.0175)\text{SlogP\_VSA2} + (0.362 \pm 0.158)\text{b\_rotN} + (0.318 \pm 0.276)\text{opr\_leadlike} - (9.717 \pm 2.221)$$

$$N_{\text{train}} = 31, R^2_{\text{train}} = 0.935, F_{\text{train}} = 72.353 > F_{0.005(5,25)} = 4.43 \text{ (the cut off value of } F \text{ distribution), } RMSE_{\text{train}} = 0.219, Q^2_{\text{LOO}} = 0.899, RMSE_{\text{LOO}} = 0.299, N_{\text{test}} = 10, R^2_{\text{test}} = 0.902, RMSE_{\text{test}} = 0.534.$$

where,  $N_{\text{train}}$  and  $N_{\text{test}}$  are the number of compounds in the training set and the test set, respectively.  $R^2_{\text{train}}$  and  $R^2_{\text{test}}$  are the squared correlation coefficient of training set and test set, respectively;  $Q^2_{\text{LOO}}$  is the leave-one-out (LOO) cross-validation squared correlation coefficient;  $F$  is the  $F$ -test value;  $RMSE$  is root mean standard error. The selected variables and their chemical meanings, standard coefficients are shown in Table 1. A variable inflation factor (VIF) ( $\text{VIF} = 1/(1 - R_j^2)$ ,  $R_j^2$  represents the multiple correlation coefficient of one descriptor's effect on the remaining molecular descriptors) was calculated

to determine if multicollinearity existed among the descriptors in models. If VIF arrays from 1.0 to 5.0, the linked equation is suitable [16]. As shown in Table 1, the VIF of all descriptors were smaller than 4, indicating that the generated model possessed statistic significance and good stability. Table 2 shows the correlation matrix of the selected descriptors. From this table, it can be seen that the linear correlation coefficient value for each pair of descriptors was smaller than 0.85, suggesting that the selected descriptors were independent, meeting the important criterion for the model selections [17]. The predicted results of the MLR model are given in Table 3 and shown in Figure 1A. As described in Table 4, obviously, the MLR model was very successfully built with statistical significance and good prediction ability. The  $R^2_{\text{train}}$  value of this model reveals that it can explain 93.5% of the variance in activity. The  $Q^2_{\text{LOO}}$  value of 0.899 was much larger than 0.5, indicating that the developed model had very good stability and predictive ability. In addition, the value of  $R^2_{\text{test}}$  for the external prediction was 0.902, showing the good prediction and generalization ability.



**Figure 1.** Plots of experimental against predicted  $pEC_{50}$  values by (A) multiple linear regression (MLR) and (B) CoMFA models.

**Table 1.** Selected descriptors of multiple linear regression.

Descriptor	Chemical Meaning	Coefficient	VIF	Stand Coefficient
b_rotN	Number of rotatable single bonds	0.362	2.888	0.408
RPC <sup>-</sup>	Relative negative partial charges	14.001	1.171	0.393
opr_leadlike	One if and only if the number of violations of Oprea's lead-like test <2 otherwise zero	0.318	1.271	0.136
SlogP_VSA2	The subdivided surface area descriptor, which is based on sum of the approximate accessible van der Waal's surface area	-0.049	3.728	-0.567
ASA	Water accessible surface area calculated using a radius of 1.4 Å for the water molecule	0.016	1.539	0.848
Constant	-	-9.717	-	-

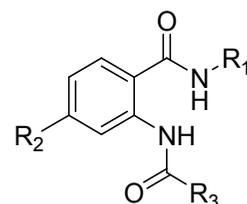
**Table 2.** The correlation matrix of descriptors.

Descriptor	b_rotN	RPC <sup>-</sup>	opr_leadlike	SlogP_VSA2	ASA
b_rotN	1.000	0.359	0.137	0.715	-0.032
RPC <sup>-</sup>	0.359	1.000	-0.010	0.215	-0.289
opr_leadlike	0.137	-0.010	1.000	0.298	-0.461
SlogP_VSA2	0.715	0.215	0.298	1.000	-0.366
ASA	-0.032	-0.289	-0.461	-0.366	1.000

**Table 3.** Molecular structures and corresponding experimental and predicted pEC<sub>50</sub> values of the AAD as partial FXR agonists.

NO.	R <sub>1</sub>	R <sub>2</sub>	R <sub>3</sub>	EC <sub>50</sub> (μM)	pEC <sub>50</sub>	2D-Pred	3D-Pred
						SW-MLR	CoMFA
1	3-carboxyphenyl	H	4- <i>tert</i> -butylphenyl	0.28	6.553	6.784	6.704
2 *	3-carboxyphenyl	H	4-(trifluoromethyl)phenyl	6.9	5.161	5.653	5.171
3	3-carboxyphenyl	H	4-bromophenyl	3.7	5.432	5.451	5.485
4	3-carboxyphenyl	H	benzo[ <i>d</i> ][1,3]dioxol-5-yl	10	5.000	5.336	5.133
5 *	3-carboxyphenyl	H	2,3-dihydrobenzo[ <i>b</i> ][1,4]dioxin-6-yl	4.9	5.310	5.714	5.133
6	3-carboxyphenyl	H	3-fluoro-4-(trifluoromethyl)phenyl	5	5.301	5.232	5.358
7 *	3-carboxyphenyl	H	styryl	5.2	5.284	4.850	5.639
8	3-acetylphenyl	H	4- <i>tert</i> -butylphenyl	0.48	6.319	6.592	6.775
9	3-cyanophenyl	H	4- <i>tert</i> -butylphenyl	0.23	6.638	6.619	6.506
10	3-methoxyphenyl	H	4- <i>tert</i> -butylphenyl	0.38	6.420	6.661	6.327
11	3-(methylthio)phenyl	H	4- <i>tert</i> -butylphenyl	0.2	6.699	6.616	6.655
12	3-(1 <i>H</i> -tetrazol-5-yl)phenyl	H	4- <i>tert</i> -butylphenyl	2.9	5.538	5.854	5.695
13	3-carbamoylphenyl	H	4- <i>tert</i> -butylphenyl	0.074	7.131	7.203	7.099
14 *	3,4-bimethoxyphenyl	H	4- <i>tert</i> -butylphenyl	0.071	7.149	7.771	6.606
15	3-(carboxymethyl)phenyl	H	4- <i>tert</i> -butylphenyl	0.42	6.377	6.493	6.414
16	3-(2-carboxyethyl)phenyl	H	4- <i>tert</i> -butylphenyl	0.064	7.194	7.143	6.999
17	2-methyl-3-carboxylphenyl	H	4- <i>tert</i> -butylphenyl	0.042	7.377	7.115	7.266
18	2-methoxy-5-carboxyphenyl	H	4- <i>tert</i> -butylphenyl	4.7	5.328	5.307	5.444
19	2-fluoro-5-carboxyphenyl	H	4- <i>tert</i> -butylphenyl	0.48	6.319	6.148	6.361
20*	2-chloro-5-carboxyphenyl	H	4- <i>tert</i> -butylphenyl	1.1	5.959	6.569	5.759
21	3-carboxy-4-methylphenyl	H	4- <i>tert</i> -butylphenyl	0.045	7.347	7.209	6.934
22 *	3-carboxy-4-methoxyphenyl	H	4- <i>tert</i> -butylphenyl	0.047	7.328	7.650	7.563
23	3-carboxy-4-chlorophenyl	H	4- <i>tert</i> -butylphenyl	0.28	6.553	6.845	6.931
24	3-carboxy-4-bromophenyl	H	4- <i>tert</i> -butylphenyl	0.15	6.824	6.937	6.889
25	3-carboxyphenyl	chloro	4- <i>tert</i> -butylphenyl	0.047	7.328	6.862	6.868
26 *	4-carboxymethylphenyl	H	naphthalen-2-yl	3.1	5.509	5.492	5.082

Table 3. Cont.



NO.	R <sub>1</sub>	R <sub>2</sub>	R <sub>3</sub>	EC <sub>50</sub> (μM)	pEC <sub>50</sub>	2D-Pred	3D-Pred
						SW-MLR	CoMFA
27	3-carboxy-4-methylphenyl	methyl	4- <i>tert</i> -butylphenyl	0.043	7.367	7.048	7.510
28	3-carboxyphenyl	methyl	4- <i>tert</i> -butylphenyl	0.061	7.215	7.334	6.942
29	3-carboxyphenyl	bromo	4- <i>tert</i> -butylphenyl	0.048	7.319	7.160	7.241
30	3-carboxyphenyl	methoxy	4- <i>tert</i> -butylphenyl	0.008	8.097	7.888	8.175
31	3-carboxy-4-methylphenyl	chloro	4- <i>tert</i> -butylphenyl	0.11	6.959	7.178	7.027
32 *	3-carboxy-4-methylphenyl	methoxy	4- <i>tert</i> -butylphenyl	0.087	7.060	7.861	6.858
33	3-carboxypropyl	H	4-ethylphenyl	5.8	5.237	5.058	5.253
34	3-carboxypropyl	H	4- <i>tert</i> -butylphenyl	2.5	5.602	5.748	5.885
35	3-carboxypropyl	H	naphthalen-2-yl	8.6	5.066	4.598	4.817
36	4-carboxybutyl	H	naphthalen-2-yl	8.3	5.081	5.086	4.795
37 *	3-methoxy-3-oxopropyl	H	naphthalen-2-yl	7.1	5.149	5.377	4.857
38	5-carboxypentyl	H	naphthalen-2-yl	4.4	5.357	5.657	5.416
39	4-carboxyphenyl	H	naphthalen-2-yl	1.0	6.000	5.882	5.902
40	3-carboxyphenyl	H	naphthalen-2-yl	1.5	5.824	5.750	5.990
41 *	4-carboxybenzyl	H	naphthalen-2-yl	1.3	5.886	6.722	5.405

\* denotes the test set compounds.

**Table 4.** Statistical parameters obtained using the MLR and CoMFA models.

Model	Training Set				Test Set		
	$R^2_{\text{train}}$	$RMSE_{\text{train}}$	$F$	$Q^2_{\text{LOO}}$	$RMSE_{\text{LOO}}$	$R^2_{\text{test}}$	$RMSE_{\text{test}}$
MLR	0.935	0.219	72.353	0.899	0.299	0.902	0.534
CoMFA-1	0.944	0.203	109.711	0.802	0.383	0.892	0.330

Finally, to confirm the robustness of the model, the Y-randomization test was performed in this study. The dependent variable vector is randomly shuffled and a new model is constructed. If the new model gives significantly lower values for both  $R^2_{\text{train}}$  and  $Q^2_{\text{LOO}}$  statistics compared to the original model, the original generated model is not considered as resulting from a chance correlation [18]. The results of ten Y-randomization tests are summarized in Table 5. As can be seen, all new  $R^2_{\text{train}}$  and  $Q^2_{\text{LOO}}$  values were much lower than those of the original model. Thereby, the good results for the MLR model were not due to a chance correlation or structural dependency of the training set.

**Table 5.**  $R^2_{\text{train}}$  and  $Q^2_{\text{LOO}}$  values after Several Y-randomization tests.

No.	1	2	3	4	5	6	7	8	9	10
$R^2_{\text{train}}$	0.150	0.145	0.188	0.08	0.123	0.179	0.144	0.186	0.141	0.131
$Q^2_{\text{LOO}}$	0.013	0.058	0.016	0.039	0.014	0.012	0.105	0.000	0.005	0.014

### 2.1.2. Model Applicability Domain Analysis for the MLR Model

Finally, to evaluate the generalization degree of the generated model, the applicability domain (AD) was defined by a Williams plot. In the Williams plot, leverage values *versus* standardized residuals were plotted to detect both the structurally influential chemicals (X outliers) and the response outliers (Y outliers) [19]. The leverage value  $h$  is defined as:

$$h_i = x_i^T (X^T X)^{-1} x_i \quad (i = 1, 2, \dots, n) \quad (1)$$

where  $x_i$  is the descriptor row vector of compound,  $X$  is the matrix of the descriptor values of the training set and  $n$  is the number of training sets. The superscript "T" refers to the transposed value of the matrix/vector [19,20]. When a leverage value  $h$  is higher than the threshold value  $h^*$  (calculated as  $3(m+1)/n$ , where  $m$  is the number of model parameters and  $n$  is the number of the training set), it is considered as X outliers. In addition, a value of  $\pm 3.0$  standard deviation units is widely used as a cut off value for defining Y outliers.

In this study, the Williams plot for the MLR model is shown in Figure 2. From this Figure 2, it can be found that no Y outliers existed in the investigated data set. Nevertheless, there were five molecules (compound 18, 20, 32, 37 and 26) with a leverage value higher than the warning leverage limit (0.581), but their predicted values were very satisfactory, with standard residuals lower than  $\pm 1.0$  standard deviation units. Hence, these molecules were not influential in the fitting performance of the model. Conversely, it further showed the reliability of the predictions of the generated model as well as confirmed its good generalization ability [19]. Therefore, compounds with high value of leverage and good fitting in the developed model can stabilize the model, and not be considered as X outliers.

As can be seen from the above results, the MLR model was significantly highly predictive, reliable and robust. It can be used to predict the FXR agonist activity of new AAD.

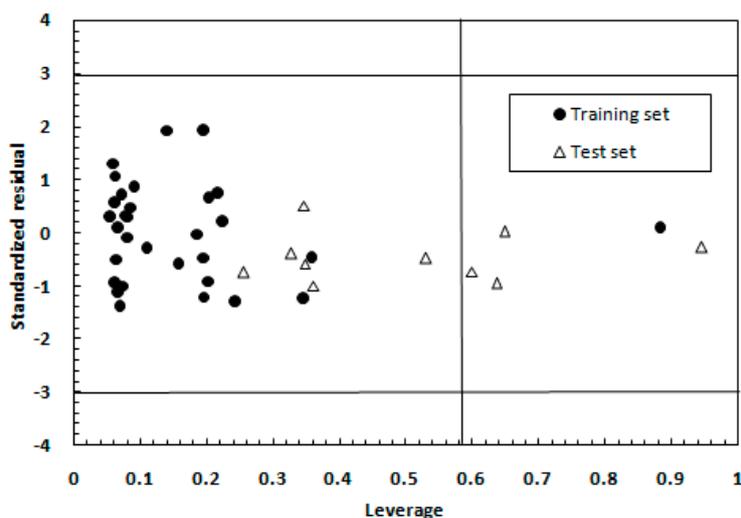


Figure 2. The Williams plot for the MLR model.

### 2.1.3. Interpretation of the Descriptors

The MLR model encompassed five descriptors:  $b_{\text{rotN}}$ ,  $\text{RPC}^-$ ,  $\text{opr\_leadlike}$ ,  $\text{SlogP\_VSA2}$  and  $\text{ASA}$ , indicating some vital physicochemical features of AAD to govern the FXR agonist activity. The relative importance of the descriptors in the model was varied in view of their standardized regression coefficients shown in Table 1 [19]. Therefore, the relative importance order is  $\text{ASA} > \text{SlogP\_VSA2} > b_{\text{rotN}} > \text{RPC}^- > \text{opr\_leadlike}$ .  $\text{ASA}$  is the water accessible surface area calculated using a radius of 1.4 Å for the water molecule. This showed that the water accessible surface area of FXR agonists might influence their agonist activity. Its positive coefficient value indicated that high polar groups tend to increase the agonist activity. For instance, it can be observed from the agonist activity of compounds 9 (having 3-cyanophenyl with  $\text{pEC}_{50} = 6.638$ ) and 10 (having 3-methoxyphenyl with  $\text{pEC}_{50} = 6.420$ ) or 1 (having 3-carboxyphenyl with  $\text{pEC}_{50} = 6.553$ ) and 8 (having acetylphenyl with  $\text{pEC}_{50} = 6.319$ ) in Table 3.  $\text{SlogP\_VSA2}$  is the subdivided surface area descriptor, which is based on the sum of the approximate accessible van der Waal's surface area, calculated for each atom with contribution to the log of the partition coefficient (octanol/water) in the range of  $(-0.2, 0)$ . Its negative coefficient value indicated that high hydrophobicity tended to decrease the agonist activity. Obviously, the bioactivity of molecules with aliphatic chains at region A in Table 3, such as 34 (with  $\text{pEC}_{50}$  of 5.602) or 35, 36, 37 and 38 (with  $\text{pEC}_{50}$  of 5.066–5.357), was lower than those without aliphatic chains such as 1 (with  $\text{pEC}_{50}$  of 6.553) or 39 and 40 (with  $\text{pEC}_{50}$  of 5.824–6.000).  $b_{\text{rotN}}$  is the number of rotatable single bonds. Its positive coefficient illustrated that more  $b_{\text{rotN}}$  was favorable to the FXR agonist activity. For instance, the bioactivity of compounds 15 and 16 or 35, 36 and 38 are varied in order: 16 (having  $b_{\text{rotN}}$  of 10) > 15 (having  $b_{\text{rotN}}$  of 9) or 38 (having  $b_{\text{rotN}}$  of 11) > 36 (having  $b_{\text{rotN}}$  of 10) > 35 (having  $b_{\text{rotN}}$  of 9).  $\text{RPC}^-$  is a relative negative partial charge descriptor that depends on the partial charge of each atom of a chemical structure. The positive sign of this descriptor illustrated that the relative negative partial charge of the molecule was favorable for the agonist activity. It can be quickly understood by comparing molecules 13 (having  $-\text{CONH}_2$  headgroup with  $\text{pEC}_{50} = 7.131$ ) and 1 (having  $-\text{COOH}$  headgroup with  $\text{pEC}_{50} = 6.553$ ).  $\text{opr\_leadlike}$  belongs to atom count and bond count descriptors that refer to the number of violations of the Oprea's lead-like test. The positive contribution of this descriptor indicated that the high value of  $\text{opr\_leadlike}$  was beneficial to the bioactivity.

Therefore, high polar groups such as the acidic headgroup together with high values of  $\text{RPC}^-$  and  $b_{\text{rotN}}$  are favorable for FXR agonist activity. Further, the aliphatic chain has a negative effect on it.

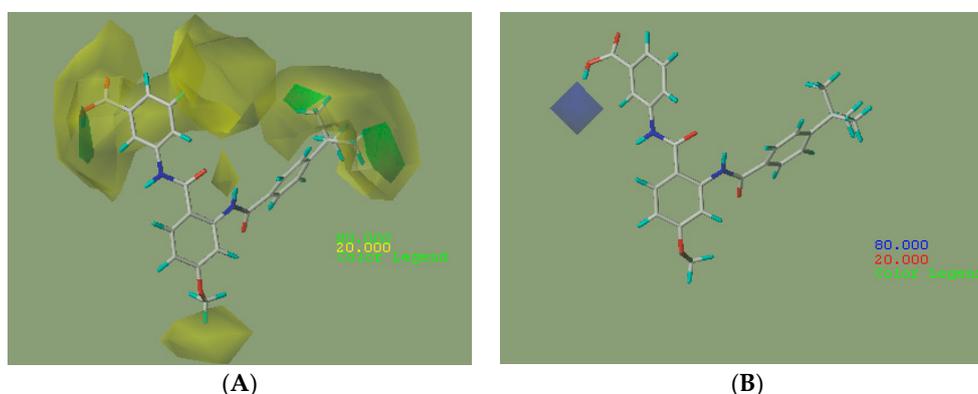
## 2.2. 3D-QSAR Models

### 2.2.1. CoMFA Analysis

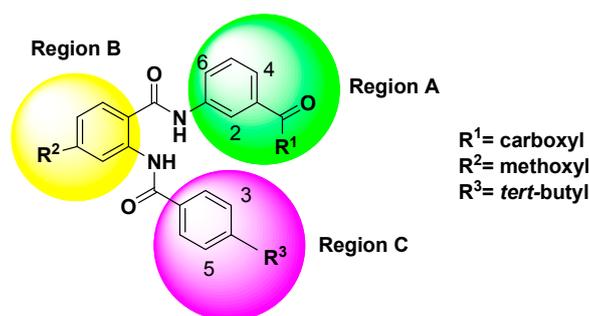
To graphically visualize the key chemical structural features that attributed to enhance the FXR agonist activity, CoMFA models were derived. The results of the CoMFA studies are listed in Table 4. The optimum number of components and filtering value for the CoMFA models were four and six, which were calculated by selecting the highest  $Q^2_{LOO}$  value. The generated CoMFA model illustrated a  $Q^2_{LOO}$  value of 0.802 ( $>0.5$ ) by four components ( $RMSE_{LOO} = 0.383$ ). The non-cross-validated PLS analysis with the four components resulted in  $R^2_{train}$  of 0.944, F value of 109.711 and  $RMSE_{train}$  of 0.203 and  $R^2_{test}$  of 0.892. The contributions of steric and electrostatic fields calculated by the CoMFA model were 49.7% and 50.3% of the variance, respectively. The obtained high  $R^2_{train}$ ,  $Q^2_{train}$  and F values along with the lower  $RMSE_{train}$  indicated the satisfactory predictive ability of the derived model (Table 4). The pEC<sub>50</sub> values predicted by the CoMFA model are listed in Table 3. Figure 1B demonstrates the correlation between experimental and predicted pEC<sub>50</sub> values by the CoMFA model.

### 2.2.2. CoMFA Contour Maps

The steric and electrostatic contour maps derived by the CoMFA model based on the reference molecule (compound 30) are shown in Figure 3. The steric interactions are represented by green and yellow contours, while electrostatic interactions are represented by red and blue contours. In the green region of the steric contour plot, bulky substitutes enhance biological activity, while in the yellow regions, these are likely to decrease the activity. Blue contours represent regions where positive charge increases activity, whereas red-colored regions represent areas where negative charge enhances activity [17]. The three regions A, B, and C of compound 30 are depicted in Figure 4.



**Figure 3.** Contour maps of the CoMFA model: (A) steric field based on compound 30; (B) electrostatic field based on compound 30. Color values specify the CoMFA field levels that enclose volumes within which increase or decrease in bulk or positive charge favor higher dependent values.



**Figure 4.** Structure of template compound (compound 30). The three regions A, B and C are depicted.

As shown in Figure 3A, there are two large yellow contours near regions A and C, one medium yellow contour near region B and one small yellow contour near region A, indicating that the bioactivity of molecules would be influenced by the introduction of bulky groups near these regions. According to Table 3, this can be explained by a comparison between molecules 21 (having  $-\text{CH}_3$  group at position 4 of region A with  $\text{pEC}_{50} = 7.347$ ) and 32 (having  $-\text{OCH}_3$  at position 4 of region A with  $\text{pEC}_{50} = 7.328$ ). The small yellow contour near position 6 of region A also suggested that the agonist activity would be decreased by the introduction of bulky groups here, such as compounds (18, 20, 19, 1) where the use of bulky groups ( $-\text{OCH}_3 > -\text{Cl} > -\text{F} > -\text{H}$ ) resulted in lower  $\text{pEC}_{50}$  values ( $5.328 < 5.959 < 6.319 < 6.553$ ). A medium yellow contour near  $\text{R}^2$  at region B indicated that bulky groups here would cause lower activity. This can be observed by the comparison of molecules 27 (substituted by  $-\text{CH}_3$  with  $\text{pEC}_{50}$  value of 7.367) and 32 (substituted by  $-\text{OCH}_3$  with  $\text{pEC}_{50}$  value of 7.060), where the volume of  $-\text{CH}_3$  was smaller than  $-\text{OCH}_3$ . This can also be observed by a comparison of compounds 25 (substituted by  $-\text{Cl}$  with  $\text{pEC}_{50}$  value of 7.328) and 29 (substituted by  $-\text{Br}$  with  $\text{pEC}_{50}$  value of 7.319). Another large yellow contour near region C showed that bulky groups at positions 3 and 5 of region C would lead to lower activity. For instance, the agonist activity of compounds 35 or 40 (substituted by naphthalen-2-yl) and 34 or 1 (substituted by 4-*tert*-butylphenyl) was varied in the order:  $35 < 34$  or  $40 < 1$ . One small green contour near positions 2 and 3 of region A indicated that FXR agonist activity would be enhanced by introduction of bulky groups here. This can be observed by comparing molecules 17 (having  $-\text{CH}_3$  group at position 2 of region A) and 1 (having  $-\text{H}$  group at position 2 of region A), where using a bulky group influenced the outcome of  $\text{pEC}_{50}$  values ( $7.377 > 6.553$ ). This can also be observed by comparing the bioactivity of molecules 16 and 15, where using bulky groups ( $-\text{CH}_2\text{CH}_2\text{COOH} > -\text{CH}_2\text{COOH}$ ) at position 3 of region A lead to higher  $\text{pEC}_{50}$  values ( $7.194 > 6.377$ ). Another two small green contours near  $\text{R}^3$  substituent groups at region C showed the favorable effect of bulky groups here in increasing the biological activity of compounds. For instance, this can be explained by comparing the activity of compounds 1 (substituted by  $-\text{C}(\text{CH}_3)_3$  with  $\text{pEC}_{50}$  value of 6.553) and 2 (substituted by  $-\text{CF}_3$  with  $\text{pEC}_{50}$  value of 5.161) or compounds 34 (substituted by  $-\text{C}(\text{CH}_3)_3$  with  $\text{pEC}_{50}$  value of 5.602) and 33 (substituted by  $-\text{CH}_2\text{CH}_3$  group with  $\text{pEC}_{50}$  value of 5.237).

As can be seen from Figure 3B, there was one medium blue contour near positions 2 and  $\text{R}^1$  of region A, which showed the favorable effect of electro-donating groups in increasing the biological activity of compounds. For instance, it can be observed by molecules 17 (having  $-\text{CH}_3$  group at position 2 of region A with  $\text{pEC}_{50}$  value of 7.377) and 1 (having  $-\text{H}$  group at position 2 of region A with  $\text{pEC}_{50}$  value of 6.553). This also can be explained by comparing the activity of compounds 13, 1 and 8, where using electro-donating substituents at  $\text{R}^1$  ( $-\text{NH}_2 > -\text{OH} > -\text{CH}_3$ ) would result in higher  $\text{pEC}_{50}$  values ( $7.131 > 6.553 > 6.319$ ).

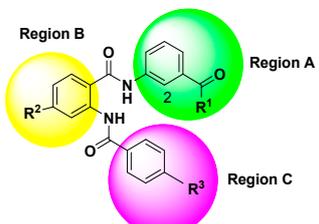
### 2.3. Design of New Partial FXR Agonists

#### 2.3.1. Chemical Structure Design

According to the information derived from the contour maps generated by the 3D-QSAR models, some important information about the chemical structures requirement was presented to investigate the effect of each kind of group as the substituent for regions A, B and C on FXR agonist activity. The bulky groups with lower electronegativity at positions 2 and  $\text{R}^1$  substituent of region A together with bulky groups at  $\text{R}^3$  of region C were considered to enhance the FXR agonist activity. However, the presence of bulky groups at positions 4 and 6 of region A,  $\text{R}^2$  of region B and 3 and 5 of region C would decrease the agonist activity. Therefore, some new compounds as potent FXR agonists were designed and are listed in Table 6. To investigate the results of each substituent on the activity results, CoMFA was the best modeling tool for use. The newly designed compounds showed that the bulky groups with lower electronegativity at  $\text{R}^1$  of region A had positive effects. This can be observed by comparing compound N3 (having  $-\text{N}(\text{CH}_3)_2$  at  $\text{R}^1$  of region A with predicted  $\text{pEC}_{50}$  value of 8.322) with template compound T30 (having  $-\text{OH}$  at  $\text{R}^1$  of region A with predicted and actual  $\text{pEC}_{50}$  values of 8.175 and

8.097, respectively). The next attempt was to improve the effects of functional groups at R<sup>2</sup> of region B where bulky effects were presented. It was observed that the addition of less bulky groups (such as -CH<sub>3</sub>) at R<sup>2</sup> of region B (Table 6) can lead to better agonist activity (compound N2 with predicted pEC<sub>50</sub> value of 8.323). Then, to investigate the bulky effects at R<sup>3</sup> of region C, different bulky groups were tried. This can be observed by comparing compounds N1, N2, N4 (having -C(CF<sub>3</sub>)<sub>3</sub>, -C(CH<sub>3</sub>)<sub>3</sub> and -Cl<sub>3</sub> at R<sup>3</sup> of region C with predicted pEC<sub>50</sub> values of 8.350, 8.323 and 8.304, respectively). Finally, to observe the effects of the addition of lower electronegativity groups at positions 2 of region A, the donor groups were investigated. It can be seen that using donor group substituent (-H < -CH<sub>3</sub> < -OH) at R<sup>2</sup> of region A would lead to an increase in the predicted pEC<sub>50</sub> values in the compounds: N7 (having -CH<sub>3</sub> with pEC<sub>50</sub> = 8.374), N9 (having -OH with pEC<sub>50</sub> = 8.388) and N1 (having -H with pEC<sub>50</sub> = 8.350). Among the designed compounds, N9 presented the highest activity with a pEC<sub>50</sub> value of 8.388. To understand the origin of this increase in activity, compounds T30 and N3 can be compared.

**Table 6.** Chemical Structures of Newly Designed partial FXR agonists based on 3D-QSAR Models.



Name	R <sup>1</sup>	Substituents at Position 2	R <sup>2</sup>	R <sup>3</sup>	Predicted pEC <sub>50</sub> Values		Docking Scores
					SW-MLR	CoMFA	
T30	OH	H	OCH <sub>3</sub>	C(CH <sub>3</sub> ) <sub>3</sub>	7.888	8.175	-10.176
N1	N(CH <sub>3</sub> ) <sub>2</sub>	H	CH <sub>3</sub>	C(CF <sub>3</sub> ) <sub>3</sub>	9.032	8.350	-14.053
N2	N(CH <sub>3</sub> ) <sub>2</sub>	H	CH <sub>3</sub>	C(CH <sub>3</sub> ) <sub>3</sub>	8.274	8.323	-11.081
N3	N(CH <sub>3</sub> ) <sub>2</sub>	H	OCH <sub>3</sub>	C(CH <sub>3</sub> ) <sub>3</sub>	8.626	8.322	-10.716
N4	N(CH <sub>3</sub> ) <sub>2</sub>	H	CH <sub>3</sub>	Cl <sub>3</sub>	8.744	8.304	-12.038
N5	N(CH <sub>3</sub> ) <sub>2</sub>	OH	CH <sub>3</sub>	C(CH <sub>3</sub> ) <sub>3</sub>	8.260	8.360	-11.602
N6	N(CH <sub>3</sub> ) <sub>2</sub>	OH	CH <sub>3</sub>	Cl <sub>3</sub>	8.828	8.357	-12.073
N7	N(CH <sub>3</sub> ) <sub>2</sub>	CH <sub>3</sub>	CH <sub>3</sub>	C(CF <sub>3</sub> ) <sub>3</sub>	9.123	8.374	-14.193
N8	N(CH <sub>3</sub> ) <sub>2</sub>	NH <sub>2</sub>	CH <sub>3</sub>	C(CF <sub>3</sub> ) <sub>3</sub>	8.816	8.378	-14.335
N9	N(CH <sub>3</sub> ) <sub>2</sub>	OH	CH <sub>3</sub>	C(CF <sub>3</sub> ) <sub>3</sub>	9.024	8.388	-14.347

R<sup>1</sup>, R<sup>2</sup>, R<sup>3</sup> represent substituent groups, respectively.

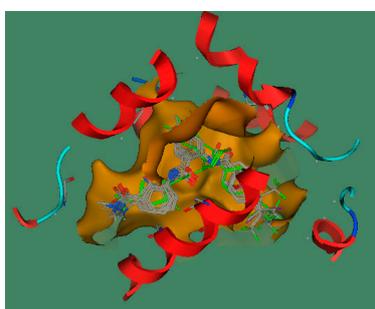
### 2.3.2. Molecular Docking Study

These molecules were ideally best based on their chemical structures, physicochemical properties and biological activity. Hence, molecular docking embedded in Molecular Operating Environment (MOE2008.10, Chemical Computing Group, Montreal, QC, Canada) was applied to study the binding modes and important interactions.

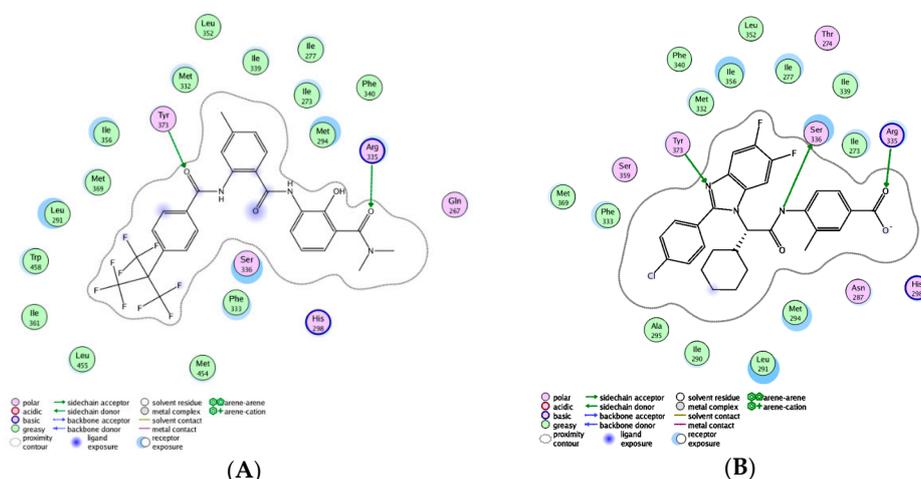
Prior to the docking, the crystal structure of FXR complexed with benzimidazole-based partial agonistic ligand was first downloaded from a protein data bank (PDB: 3OLF). The protein was protonated using the AMBER99 force field. A set of possible conformations of nine newly designed molecules was prepared by the conformational generation function of MOE. Then, molecular docking was carried out by following parameters: the binding site was defined by the ligand atom mode; triangle matcher was used as a placement method; two rescoring methods were computed, rescoring 1 was selected as London dG; rescoring 2 was selected as affinity; force field was used as a refinement [21].

A critical factor that determines the effectiveness of a docking program is its ability to reproduce ligand poses in the receptor as close to those found in X-ray deduced structures as possible [22]. In this docking study, the root-mean-square distance (RMSD) parameter measured between the complexed ligand and the redocked ligand was 0.5749 Å, suggesting that the docking results were very suitable

and reliable. Docking results are listed in Table 6. Obviously, these newly designed compounds had higher docking scores for FXR than the known template compound T30, which was in agreement with the 2D- and 3D-QSAR results. The best docked orientation of compounds is shown in Figure 5, showing that newly designed compounds can be well docked into the ligand binding site of FXR. The best docked conformation of the most active compound N9, as shown in Figure 6A, revealed that the presence of perfluoroalkyl chain substituted groups at region C allowed for potentiation of strong hydrophobic interactions with Met369, Leu291, Trp458, Met454, Ile361, Leu455 and Phe333 in the active site of FXR and formed two H-bonds with Arg335 and Tyr373. Comparative molecular docking between compound N9 and the complexed ligand, shown in Figure 6, indicated that the former had a better binding score than the latter, suggesting that hydrophobic interactions between groups at region C with these amino acids played a dominant role in the FXR agonist activity. Thereby, the hydrophobic interaction of groups at region C seems to stabilize the compound within the binding site, thus contributing greater activity.



**Figure 5.** The best docked conformations and poses of newly designed compounds in the ligand binding domain of FXR.



**Figure 6.** The 2D representation of docking of compounds N9 (A) and complexed ligand (B) into the FXR active site.

### 3. Experimental Section

#### 3.1. Data Set

The structures and biological activities of 41 AAD as FXR agonists used for the QSAR analyses were taken from Merk *et al.* [8,9] and are listed in Table 3. The agonist activity ( $EC_{50}$ ) value was converted to a logarithmic-scale  $pEC_{50}$  value, which was taken as the dependent parameter for the QSAR study. In order to establish a reliable model, the data set was randomly divided into two subsets,

a training set of 31 compounds (approximately 75% of the data) that represented a wide range of varied structures and a test set of 10 compounds (approximately 25%) that followed the distribution of the activity values for the training set [14]. The training set is to build models, while the test set marked by the asterisk in Table 3 will be used to evaluate the prediction ability of the final training set model.

### 3.2. 2D-QSAR Studies

#### 3.2.1. Descriptors Calculation

All 2D structures of the molecules in Table 3 were sketched and their 3D structures were subjected to energy minimization using the molecular mechanics force field (MMFF) method with a convergence criterion of 0.01 kcal/mol and partial atomic charges. The final geometry optimization of each energy-minimized structure was carried out by stochastic conformational search. Then, only the lowest energy conformer of each structure was used to calculate 327 descriptors by employing the QuaSAR module of MOE [23]. These calculated descriptors include three classes: 2D descriptors, which use the atoms and connection information of the molecules, internal 3D (i3D), which uses 3D coordinate information about each molecule and external 3D (x3D), which uses 3D coordinate information with an absolute frame of reference. All the above processes were performed using MOE2008.10 package.

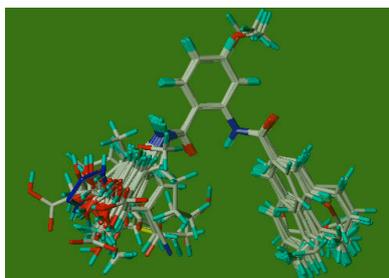
#### 3.2.2. Stepwise Multiple Linear Regression (SW-MLR)

In this study, a stepwise technology combined with MLR (SW-MLR) was employed to select a set of the most relevant descriptors for model constructions. The stepwise regression combines forward and backward selections. It selects statistically meaningful variables that can appreciably increase the residual sum of squares checked by the Fisher test [24]. Therefore, different MLR models will be derived in this procedure. The selection of a good MLR equation is made by statistical parameters such as the squared correlation coefficient ( $R^2$ ), root mean standard error (RMSE), and Fisher statistic [25]. The best MLR model should have high  $R^2$  and Fisher statistic, and low RMSE.

### 3.3. 3D-QSAR Studies

#### 3.3.1. Molecular Alignment

The 3D-QSAR model was constructed by CoMFA embedded in SYBYL 6.9. Because the prediction accuracy of CoMFA is highly dependent on the structural alignment of the molecules with a reference compound, the selection of the template molecule plays an important role in performing CoMFA. Generally, the lowest energy conformer of the most active compound is selected as a template molecule for superimposition of all other compounds [26]. Therefore, compound 30 (as shown in Figure 4) was identified as a reference molecule in view of its highest activity, and all of the remaining compounds were aligned on it to derive CoMFA models. The structures of the aligned molecules are demonstrated in Figure 7.



**Figure 7.** Alignment of training and test set compounds on compound 30. Baby blue, red, blue, green, gray and yellow signify hydrogen atom, oxygen atom, nitrogen atom, fluorine atom, carbon atom and sulfur atom, respectively.

### 3.3.2. CoMFA Modeling

After aligning the molecules within the lattice that extended 4 Å units beyond the align molecules in all directions, an  $sp^3$  hybridized carbon was utilized as a probe atom to generate the steric and electrostatic fields with a charge of +1.0 and a van der Waals radius of 1.52 Å. The steric and electrostatic contributions were set as a default cut-off energy value of 30 kcal/mol. A partial least-squares (PLS) method, an extension of multiple regression analysis, was applied to calculate the minimal set of grid points and then linearly correlate the CoMFA fields to the  $pEC_{50}$  values in order to generate the CoMFA model [27].

### 3.4. Model Validation

The predictive ability and reliability of 2D- and 3D-QSAR models were evaluated by internal and external validations. The leave-one-out (LOO) cross-validation technology is often considered as the best way to internally validate the quality of derived models [28]. The LOO produces a number of models by deleting one object from the training set, which employs all the information available. Generally, when the value of LOO cross validated correlation coefficient ( $Q^2_{LOO}$ ) goes over a threshold value of 0.5, the model is acceptable [29]. In addition, external validation is also essential and significant to evaluate the generalization performance of the proposed model [25,30]. The statistical parameters, such as the root mean square errors ( $RMSE_{test}$ ) and the squared correlation coefficient ( $R^2_{test}$ ) of the external test set were calculated to assess the performance of the model [31].

All algorithms were written in MATLAB 8.0 and run on a computer [Intel(R) Pentium(R), 2.00-GB RA].

## 4. Conclusions

In this paper, a three level *in silico* approach was applied to investigate some important structural and physicochemical aspects of highly potent partial FXR agonists. Initially, 2D-QSAR using methods of both SW-MLR and 3D-QSAR CoMFA studies was performed based on forty-one AAD. Satisfactory results were obtained with the proposed methods. The best derived 2D-QSAR model by SW-MLR can explain 93.5% of the variance in activity with a low RMSE of 0.219. In addition, the 2D-QSAR study demonstrated that  $b\_rotN$ ,  $RPC^-$ ,  $opr\_leadlike$ ,  $SlogP\_VSA2$ ,  $ASA$  of molecules had high correlation with the FXR agonist activity. Meanwhile, the best 3D-QSAR model presented higher predictive ability ( $R^2_{train} = 0.944$ ,  $RMSE_{train} = 0.203$ ,  $Q^2_{LOO} = 0.802$ ,  $R^2_{test} = 0.892$ ) compared with the 2D-QSAR models. The derived contour maps from the 3D-QSAR model suggested the significant structural features (steric and electronic effects) required for improving biological activity. Consequently, the bulky groups with lower electronegativity at positions 2 and  $R^1$  substituent of region A together with bulky groups at  $R^3$  of region C were considered to enhance the FXR agonist activity. However, the presence of bulky groups at positions 4 and 6 of region A,  $R^2$  of region B and 3 and 5 of region C would decrease the agonist activity. Therefore, the obtained 2D- and 3D-QSAR models could provide valuable guidance for future design of new potent partial FXR agonists with an anthranilic acid skeleton in the drug discovery process. Finally, nine newly designed AAD with predicted  $pEC_{50}$  values higher than the known template compound were docked to the ligand binding domain of FXR. The molecular binding patterns and docking scores of these nine molecules also suggested that they can be robust and potent partial FXR agonists in agreement with the QSAR results. This also revealed that the hydrophobic interaction of groups at region C with Met369, Leu291, Trp458, Met454, Ile361, Leu455 and Phe333 seemed to stabilize the compound within the binding site, thus contributing greater activity. To the best of our knowledge, this work constituted the first *in silico* study for AAD as partial FXR agonists.

**Acknowledgments:** The authors thank Ji Zhiliang for his technological help. This work is supported by National Natural Science Foundation program (81503497), Fujian Provincial Natural Science fund subject (2015J01340), the key discipline special program of Fujian university of traditional Chinese medicine (X2014012), Fujian provincial traditional Chinese medicine science and technology program (wzrk201303) and Fujian Education Department class A science and technology fund (JA14155).

**Author Contributions:** Meimei Chen contributed to the analysis of the study and manuscript writing; Xuemei Yang, Xinmei Lai, Jie Kang, Huijuan Gan and Yuxing Gao helped perform the analysis.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Düfer, M.; Hörth, K.; Wagner, R.; Schittenhelm, B.; Prowald, S.; Wagner, T.F.J.; Oberwinkler, J.; Lukowski, R.; Gonzalez, F.J.; Krippeit-Drews, P.; *et al.* Bile acids acutely stimulate insulin secretion of mouse-cells via farnesoid X receptor activation and KATP channel inhibition. *Diabetes* **2012**, *6*, 1479–1489. [[CrossRef](#)] [[PubMed](#)]
2. Nijmeijer, R.M.; Gadaleta, R.M.; van Mil, S.W.C.; van Bodegraven, A.A.; Crusius, J.B.; Dijkstra, G.; Hommes, D.W.; de Jong, D.J.; Stokkers, P.C.F.; Verspaget, H.W.; *et al.* Farnesoid X receptor (FXR) activation and FXR genetic variation in inflammatory bowel disease. *PLoS ONE* **2011**, *8*, e23745. [[CrossRef](#)] [[PubMed](#)]
3. Hollman, D.A.A.; Milona, A.; van Erpecum, K.J.; van Mil, S.W.C. Anti-inflammatory and metabolic actions of FXR: Insights into molecular mechanisms. *Biochim. Biophys. Acta* **2012**, *11*, 1443–1452. [[CrossRef](#)] [[PubMed](#)]
4. Akwabi-Ameyaw, A.; Bass, J.Y.; Caldwell, R.D.; Caravella, J.A.; Chen, L. FXR agonist activity of conformationally constrained analogs of GW4064. *Bioorg. Med. Chem. Lett.* **2009**, *19*, 4733–4739. [[CrossRef](#)] [[PubMed](#)]
5. Merk, D.; Steinhilber, D.; Schubert-Zsilavec, M. Characterizing ligands for farnesoid X receptor available *in vitro* test systems for farnesoid X receptor modulator development. *Expert Opin. Drug Discov.* **2014**, *1*, 27–37. [[CrossRef](#)] [[PubMed](#)]
6. Singh, N.; Yadav, M.; Singh, A.K.; Kumar, H.; Dwivedi, S.K. Synthetic FXR agonist GW4064 is a modulator of multiple G protein-coupled receptors. *Mol. Endocrinol.* **2014**, *5*, 659–673. [[CrossRef](#)] [[PubMed](#)]
7. Flatt, B.; Martin, R.; Wang, T.L.; Mahaney, P.; Murphy, B. Discovery of XL335 (WAY-362450), a highly potent, selective, and orally active agonist of the farnesoid X receptor (FXR). *J. Med. Chem.* **2009**, *4*, 904–907. [[CrossRef](#)] [[PubMed](#)]
8. Merk, D.; Gabler, M.; Gomez, R.C.; Flesch, D.; Hanke, T.; Kaiser, A. Anthranilic acid derivatives as novel ligands for farnesoid X receptor (FXR). *Bioorg. Med. Chem.* **2014**, *22*, 2447–2460. [[CrossRef](#)] [[PubMed](#)]
9. Merk, D.; Lamers, C.; Ahmad, K.; Gomez, R.C.; Schneider, G. Extending the structure-activity relationship of anthranilic acid derivatives as farnesoid X receptor modulators: Development of a highly potent partial farnesoid X receptor agonist. *J. Med. Chem.* **2014**, *57*, 8035–8055. [[CrossRef](#)] [[PubMed](#)]
10. Putz, M.V.; Lacrămă, A.M. Introducing spectral structure activity relationship (S-SAR) analysis. Application to ecotoxicology. *Int. J. Mol. Sci.* **2007**, *8*, 363–391. [[CrossRef](#)]
11. Putz, M.V.; Putz, A.M.; Lazea, M.; Ienciu, L.; Chiriac, A. Quantum-SAR extension of the spectral-SAR algorithm. Application to polyphenolic anticancer bioactivity. *Int. J. Mol. Sci.* **2009**, *10*, 1193–1214. [[CrossRef](#)] [[PubMed](#)]
12. Davis, G.D.J.; Vasanthi, A.H.R. QSAR based docking studies of marine algal anticancer compounds as inhibitors of protein kinase B (PKBb). *Eur. J. Pharm. Sci.* **2015**, *76*, 110–118. [[CrossRef](#)] [[PubMed](#)]
13. Noguera, G.J.; Fabian, L.E.; Lombardo, E.; Finkielstein, L. QSAR study and conformational analysis of 4-arylthiazolylhydrazones derived from 1-indanones with anti-Trypanosoma cruzi activity. *Eur. J. Pharm. Sci.* **2015**, *78*, 190–197. [[CrossRef](#)] [[PubMed](#)]
14. Roy, K.; Paul, S. Docking and 3D-QSAR studies of acetohydroxy acid synthase inhibitor sulfonylurea derivatives. *J. Mol. Model.* **2010**, *16*, 951–964. [[CrossRef](#)] [[PubMed](#)]
15. Putz, M.V.; Dudas, N.A.; Isvoran, A. Double variational binding—(SMILES) conformational analysis by docking mechanisms for anti-HIV pyrimidine ligands. *Int. J. Mol. Sci.* **2015**, *16*, 19553–19601. [[CrossRef](#)] [[PubMed](#)]
16. Goodarzi, M.; Jensen, R.; Heyden, Y.V. QSRR modeling for diverse drugs using different feature selection methods coupled with linear and nonlinear regressions. *J. Chromatogr. B* **2012**, *910*, 84–94. [[CrossRef](#)] [[PubMed](#)]
17. Jalali-Heravi, M.; Asadollahi-Baboli, M.; Shahbazikhah, P. QSAR study of heparanase inhibitors activity using artificial neural networks and Levenberge-Marquardt algorithm. *Eur. J. Med. Chem.* **2008**, *43*, 548–556. [[CrossRef](#)] [[PubMed](#)]

18. Zhao, L.; Xiang, Y.; Song, J.; Zhang, Z. A novel two-step QSAR modeling work flow to predict selectivity and activity of HDAC inhibitors. *Bioorg. Med. Chem. Lett.* **2013**, *23*, 929–933. [[CrossRef](#)] [[PubMed](#)]
19. Wei, Y.; Xi, L.; Yao, X.; Li, J.Z.; Wu, X. Quantitative structure-activity relationship analysis of a series of human renal organic anion transporter inhibitors. *Arch. Pharm. Chem. Life Sci.* **2012**, *345*, 759–766. [[CrossRef](#)] [[PubMed](#)]
20. Masand, V.H.; Mahajan, D.T.; Alafeefy, A.M.; Bukhari, S.N.A.; Elsayed, N.N. Optimization of antiproliferative activity of substituted phenyl 4-(2-oxoimidazolidin-1-yl) benzenesulfonates: QSAR and CoMFA analyses. *Eur. J. Pharm. Sci.* **2015**, *77*, 230–237. [[CrossRef](#)] [[PubMed](#)]
21. Khedr, M.A.; Shehata, T.M.; Mohamed, M.E. Repositioning of 2,4-dichlorophenoxy acetic acid as a potential anti-inflammatory agent: *In silico* and pharmaceutical formulation study. *Eur. J. Pharm. Sci.* **2014**, *65*, 130–138. [[CrossRef](#)] [[PubMed](#)]
22. Makhuri, F.R.; Ghasemi, J.B. Computer-aided scaffold hopping to identify a novel series of casein kinase 1  $\delta$  (CK1d) inhibitors for amyotrophic lateral sclerosis. *Eur. J. Pharm. Sci.* **2015**, *78*, 151–162. [[CrossRef](#)] [[PubMed](#)]
23. Chen, M.M.; Lai, X.M.; Yang, X.M. A QSAR classification study on inhibitory activities of 2-arylbenzoxazoles against cholesteryl ester transfer protein. *Med. Chem. Res.* **2014**, *23*, 1878–1886. [[CrossRef](#)]
24. Saghalei, L.; Shahlaei, M.; Fassihi, A.; Madadkar-Sobhani, A.; Gholivand, M.B.; Pourhossein, A. QSAR analysis for some diaryl-substituted pyrazoles as CCR2 Inhibitors by GA-stepwise MLR. *Chem. Biol. Drug Des.* **2011**, *77*, 75–85. [[CrossRef](#)] [[PubMed](#)]
25. Vrontaki, E.; Melagraki, G.; Mavromoustakos, T.; Afantitis, A. Exploiting ChEMBL database to identify indole analogues as HCV replication inhibitors. *Methods* **2015**, *71*, 4–13. [[CrossRef](#)] [[PubMed](#)]
26. Fortin, S.; Wei, L.; Moreau, E.; Lacroix, J.; Cote, M.F.; Petitclerc, E.; Kotra, L.P.; Gaudreault, R.C. Substituted phenyl 4-(2-oxoimidazolidin-1-yl)benzenesulfonamides as antimitotics. Antiproliferative, antiangiogenic and antitumoral activity, and quantitative structure-activity relationships. *Eur. J. Med. Chem.* **2011**, *46*, 5327–5342. [[CrossRef](#)] [[PubMed](#)]
27. Wang, Y.; Wu, M.; Ai, C.; Wang, Y. Insight into the structural determinants of imidazole scaffold-based derivatives as TNF- $\alpha$  release inhibitors by *in silico* explorations. *Int. J. Mol. Sci.* **2015**, *16*, 20118–20138. [[CrossRef](#)] [[PubMed](#)]
28. Zhang, J.; Han, B.; Wei, X.; Tan, C.; Chen, Y.; Jiang, Y. A two-step target binding and selectivity support vector machines approach for virtual screening of dopamine receptor subtype-selective ligands. *PLoS ONE* **2012**, *7*, e39076. [[CrossRef](#)] [[PubMed](#)]
29. Wegner, J.K.; Frohlich, H.; Zell, A. Feature selection for descriptor based classification models. Theory and GA-SEC algorithm. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 921–930. [[CrossRef](#)] [[PubMed](#)]
30. Golbraikh, A.; Tropsha, A. Beware of q<sup>2</sup>. *J. Mol. Graph. Model.* **2002**, *20*, 269–276. [[CrossRef](#)]
31. Shahlaei, M.; Fassihi, A. QSAR analysis of some 1-(3,3-diphenylpropyl)-piperidiny amides and ureas as CCR5 inhibitors using genetic algorithm-least square support vector machine. *Med. Chem. Res.* **2013**, *22*, 4384–4400. [[CrossRef](#)]

