Article

In Silico Searching for Alternative Lead Compounds to Treat Type 2 Diabetes through a QSAR and Molecular Dynamics Study

Nicolás Cabrera [1], Sebastián A. Cuesta [2,3], José R. Mora [3,*], Luis Calle [4,5], Edgar A. Márquez [6,*], Roland Kaunas [1] and José Luis Paz [7]

1 Department of Biomedical Engineering, Texas A&M University, College Station, TX 77843, USA; nicolascabrera93@tamu.edu (N.C.); rkaunas@tamu.edu (R.K.)
2 Department of Chemistry, Manchester Institute of Biotechnology, The University of Manchester, 131 Princess Street, Manchester M1 7DN, UK; sebastian.cuesta@postgrad.manchester.ac.uk
3 Grupo de Química Computacional y Teórica (QCT-USFQ), Departamento de Ingeniería Química, Universidad San Francisco de Quito, Diego de Robles y vía Interoceánica, Quito 170901, Ecuador
4 Faculty of Pharmacy, University of Granada, 18011 Granada, Spain; lucho_calle_mendoza@hotmail.com
5 Facultad de Ciencias Médicas, Instituto de Investigación e Innovación en Salud Integral, Universidad Católica Santiago de Guayaquil, Guayaquil 09013493, Ecuador
6 Grupo de Investigaciones en Química y Biología, Departamento de Química y Biología, Facultad de Ciencias Exactas, Universidad del Norte, Carrera 51B, Km 5, vía Puerto Colombia, Barranquilla 081007, Colombia
7 Departamento Académico de Química Inorgánica, Facultad de Química e Ingeniería Química, Universidad Nacional Mayor de San Marcos, Cercado de Lima 15081, Peru; jpazr@unmsm.edu.pe
* Correspondence: jrmora@usfq.edu.ec (J.R.M.); ebrazon@uninorte.edu.co (E.A.M.)

Abstract: Free fatty acid receptor 1 (FFA1) stimulates insulin secretion in pancreatic β-cells. An advantage of therapies that target FFA1 is their reduced risk of hypoglycemia relative to common type 2 diabetes treatments. In this work, quantitative structure–activity relationship (QSAR) approach was used to construct models to identify possible FFA1 agonists by applying four different machine-learning algorithms. The best model (M2) meets the Tropsha's test requirements and has the statistics parameters $R^2 = 0.843$, $Q^2_{CV} = 0.785$, and $Q^2_{ext} = 0.855$. Also, coverage of 100% of the test set based on the applicability domain analysis was obtained. Furthermore, a deep analysis based on the ADME predictions, molecular docking, and molecular dynamics simulations was performed. The lipophilicity and the residue interactions were used as relevant criteria for selecting a candidate from the screening of the DiaNat and DrugBank databases. Finally, the FDA-approved drugs bilastine, bromfenac, and fenofibric acid are suggested as potential and lead FFA1 agonists.

Keywords: free fatty acid receptor 1; type 2 diabetes; molecular dynamics; molecular docking; agonits of FFA1

1. Introduction

Diabetes mellitus affected approximately 463 million people worldwide (~9.3% of the population) in 2019, with numbers expecting to rise to 578 million (~10%) by 2030 and 700 million (~11%) by 2045 [1]. Type 2 Diabetes Mellitus (T2DM) is the most common type of diabetes, representing 80% of all the cases [2]. T2DM is characterized by reduced secretion of insulin from β-cells and resulting hyperglycemia associated with cardiovascular complications such as cardiac ischemia and stroke [3,4]. Current methods to control blood sugar, including controlled diet, metformin, sulfonylurea, and insulin [5], have limited efficacy and are associated with potential health problems such as hypoglycemia, weight gain, and lack of sustained efficacy [6–8].

Free Fatty Acid Receptor 1 (FFA1), a protein expressed in pancreatic β-cell, has become a target of interest since FFA1 activation through ligand-receptor binding induces insulin

secretion [7]. Unlike specific diet, metformin, and sulfonylurea, induction of insulin in response to FFA1 agonists is attenuated when blood glucose levels are excessive, providing negative feedback that reduces the risk of hypoglycemia [9].

The orthosteric drug TAK-875 is the most explored FFA1 agonist, showing a potent antidiabetic effect in early-stage clinical trials, with a lower propensity to cause hypoglycemia [10]. Unfortunately, the development of TAK-875 treatment was terminated in phase III clinical trials due to its hepatotoxicity [11]. Other FFA1 agonists have high lipophilicity, leading to poor pharmacokinetics, metabolic instability, toxicity, and off-target effects [12–16]. In 2008, Christiansen et al. discovered a series of 4-phenethynyldihydrocinnamic acid FFA1 agonists, the most potent being TUG-424; however, TUG-424 was too lipophilic to be a viable drug candidate [17]. This group subsequently synthesized several FFA1 agonists with relatively low lipophilicity, inspired by the molecular structure of TUG-424 and TAK-875 with varying levels of agonist activity [18–20]. While these studies provide important information, they lack a systematic analysis to relate agonist characteristics with their potency.

In silico analysis is an economic strategy for the exploration of new compounds [21,22]. Quantitative structure–activity relationship (QSAR) reveals relationships between structural descriptors and biological activities of chemical compounds. In this regard, a report of a multiple linear regression QSAR model to improve the FFA1 agonist activity was reported [23]. The biological activity was used as the logarithm of the half-maximal effective concentration (pEC_{50}), and the best model was statistically well-validated in terms of leave-one-out cross validation. Furthermore, the X-ray crystallographic structure of the binding mode in the FFA1 was elucidated [24], and molecular dynamics simulation [8] and experimental studies [25] clarify the activation mechanism of FFA1 to treat diabetes. Strong hydrogen bond interactions with Tyr-91, Arg-183, Asn-244, and Arg-258 are linked with higher agonist potency on the activation of FFA1 [26]. Also, molecular dynamics were used to get insights in selectivity [27,28] and allosteric activation [29,30] to design new molecules to treat T2DM. These reports lay out valuable information to be used for in silico studies to find a possible FFA1 agonist.

In this article, new FFA1 agonists are proposed based on an in silico analysis. Lipophilicity and interaction with FFA1 of the suggested drugs were studied using QSAR, molecular docking, and molecular dynamics simulation. In this sense, a diverse dataset of 93 FFA1 agonists reported by Christiansen et al. [17–20] was used to build a model. The pIC_{50} of the molecules of the dataset, DiaNat [27], and DrugBank 5.1.7 (https://www.drugbank.ca/ (accessed on 18 February 2021)) [28] databases were predicted. In addition, a computational analysis of absorption, distribution, metabolism, and excretion (ADME) parameters was performed to explore lipophilicity, the effectiveness of the drug, and the affinity to reach the target site [29]. Finally, the crystal structure of FFA1 was used (PDB: 4PHU) for molecular docking and molecular dynamic simulations [30,31].

2. Materials and Methods

The step-by-step process of this study is shown in Scheme 1.
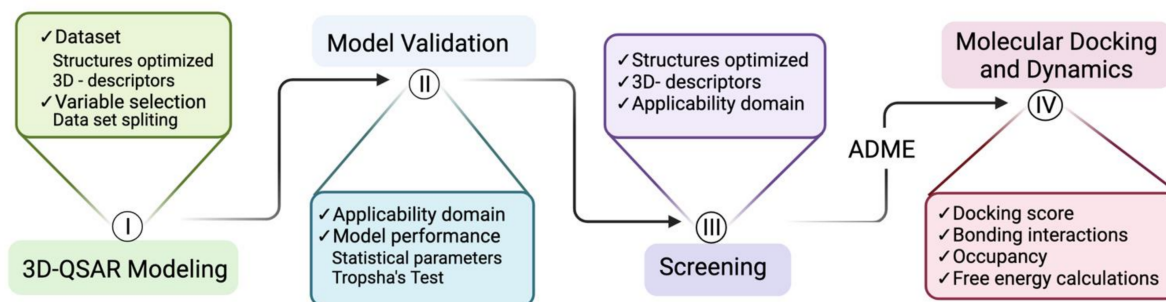
Scheme 1. Flow chart of the step-by-step process applied in this study (Created with BioRender.com).

2.1. Dataset

The molecular structure and experimental pEC_{50} values of 93 FFA1 agonists were collected from reports by Christiansen et al. [32]. The 3D structures of these molecules were further optimized based on Universal Force Field (UFF) molecular mechanics theory [32] with RDKit software (https://www.rdkit.org/ (accessed on 18 February 2021)) implemented in QuBiLs-MIDAS software v2.0. The simplified molecular-input line-entry system (SMILES) and pEC_{50} values of each compound are available in supporting information (Table S1). Topographic 3D molecular descriptors suggested for cheminformatics studies such as QSAR were calculated using QuBiLs-MIDAS [33]. A total of 3031 3D molecular descriptors were estimated based on many algebraic operations that consider linear, bilinear, and quadratic indexes for all molecules in the dataset [34]. Next, two subsets were obtained for the modelling process. The first subset with 1213 descriptors (SS_1213) was selected by applying Shannon entropy of 0.7 and Spearman coefficient of 0.7 as cutoffs. The Shannon entropy cutoff was used to identify descriptors to be removed when more than 30% of compounds have the same descriptor value, and the Spearman coefficient cutoff establishes the maximum allowable correlation between two descriptors. The second subset with 1393 descriptors (CSE_1393) was obtained by applying the attribute evaluator CfsSubsetEval implemented in Weka 3.8 [35], which evaluates the predictive ability of each descriptor in a supervised way for the response variable (i.e., pEC_{50}). The compounds in the dataset were divided into training (75%) and test (25%) based on the Ward's method of clustering to minimize the error of the sum of squares within clusters [36]; Ward's method is a hierarchical cluster analysis for a rational split of the molecules into training and test set [37].

2.2. Variable Selection

Variable selection for the modelling process avoids overfitting, redundancy, and irrelevancy when models are constructed with few descriptors [38]. Variable selection was performed on Weka 3.8 with the wrapper method and the following machine learning techniques: multilinear regression (MLR), random forest, instance-based learning with *parameter k* (IBK), and Smola and Scholkopf's algorithm for solving regression problem (SMOreg) [39]. The wrapper method uses a classifier to find a good set of descriptors by searching through the descriptor space and wraps a classifier in a cross-validation loop [40]. The aforementioned machine-learning techniques are more accurate than filter methods that evaluate the relevance of features based on high or low scores [41]. Then, the group of selected descriptors was assessed as possible models based on the statistical parameters obtained.

2.3. Applicability Domain

Applicability domain (AD) analysis is commonly performed to increase the confidence and reliability in predictions on QSAR models and is now considered a requirement for this type of modelling [42]. AD is the physicochemical, structural, biological space, or information on which the training set was developed and defines the space for reliable prediction of new drugs [43]. The AD in the current study was defined by the consensus strategy suggested in AMBIT discovery (http://ambit.sourceforge.net/ (accessed on 18 February 2021)). The AD methods used for the consensus are principal component analysis (PCA), range-based, probability density, Euclidean, and city block distance. The consensus score determines if a compound is inside or outside of the AD [44]. If three or more methods consider a compound as an outlier (score ≤ 0.25), that molecule is excluded from further dataset analysis.

2.4. Model Performance

Model performance was validated using well-known statistical parameters: coefficient of determination (R^2), 5-fold cross-validation coefficient (Q^2_{CV}), the external validation coefficient (Q^2_{ext}), bootstrapping coefficient (Q^2_{boot}), Y-scrambling analysis, and Tropsha's

test [45] (available at www.oecd.org (accessed on 18 February 2021)). The functions of these statistical parameters are shown in Table 1. Also, the collinearity between the descriptors expressed as the Pearson's correlation coefficient was evaluated.

Table 1. Statistical parameters to evaluate the performance of the models.

Parameter	Function
R^2	Global evaluator of the model.
Q^2_{CV}, and Q^2_{ext}	Evaluators of the predictability of the model.
(Q^2_{boot})	Evaluate the randomly generating training sets with 5000 sample repetition, and the predicted response of each sample is obtained.
Y-scrambling analysis	A random perturbation of the response variable.

2.5. Screening of the DrugBank 5.1.7 and DiaNat Databases

The QSAR model was employed to screen the DiaNat [27] and DrugBank 5.1.7 [28] databases in order to identify possible FFA1 agonists. DiaNat is composed of 336 natural products extracted from different medicinal plants. The antidiabetic activity of these natural products was tested either in vivo or in vitro. DrugBank 5.1.7 comprises 2636 approved drugs, 6127 experimental compounds, 118 nutraceutical compounds, and 245 withdrawn compounds. The candidates' molecular structures were evaluated in detail and compared with dataset's molecule structures.

2.6. Absorption, Distribution, Metabolism, and Excretion (ADME) Predictions

ADME parameters are routinely used to analyze the drug-likeliness and pharmacokinetics of potential drugs. In this study, SwissADME (http://www.swissadme.ch (accessed on 18 February 2021)) was used to predict ADME parameters for the identified FFA1 agonists. SwissADME canonicalizes, processes the SMILES of each molecule, adding hydrogens, neutralizing, and obtaining the Kekulé's 3D structure.

Lipophilicity, the drug's affinity for a lipid environment, is a key property in the design of possible FFA1 agonists, since high lipophilicity renders a drug candidate unsuitable [17]. Lipophilicity is typically expressed by the partition coefficient between *n*-octanol and water (log $P_{o/w}$). A consensus log $P_{o/w}$ was computed from the average values of five lipophilicity parameters: iLogP [46], XLogP3 [47], WLogP [48], MLogP [49], and SILICOS-IT [50]. The lipophilicity value obtained from an atomistic method on the fragmental system of Wildman and Crippen (WLOGP) and the topological polar surface area (TPSA) were used to construct the BOILED-Egg depictive model, which shows the drug probability of human intestinal absorption (HIA) and blood-brain barrier (BBB) permeation [51].

2.7. Molecular Docking

To gain insight into how the studied compounds fit into the target pocket and which amino acids are involved in the receptor–ligand interactions, a molecular docking analysis was performed targeting the orthosteric site of FFA1 based on its elucidated three-dimensional X-ray crystallographic structure (PDB:4PHU) [24]. The FFA1 structure was prepared by removing water molecules, ligands, and other external residues using Pymol [52]. AutoDockTools [53] was used to add polar hydrogens and rotatable bonds in the ligand. The centered coordinates for the grid box were obtained from the retrieved protein using Discovery Studio Visualizer V20.1.0 [54] (Xc = −50.057, Yc = −2.660, Zc = 57.856). Docking calculations were performed using AutoDock Vina using a grid box size of 24 × 18 × 18 (X, Y, Z, respectively), with default exhaustiveness, full ligand flexibility, and a 1 Å spacing [55].

2.8. Molecular Dynamics

Molecular dynamics simulations were performed on the most promising agonist candidates to gain insight into the protein–ligand interactions, including hydrogen bond formation and evaluating the stability of ligand-binding within the orthostatic site of FFA1. The conformations obtained from the docking calculation were used as the starting point to

run the dynamic simulations using Gromacs 2019 [56]. First, the protein topology was built using AMBER99SB-ILDN [57] implemented in Gromacs 2019. Next, the ligands topology was built using antechamber python parser interface (ACPYPE) server and the Generalized Amber Force Field (GAFF) [58]. Finally, the system was solvated in a cube shape using the three-point water model (TIP3) and neutralized by adding sodium or chlorine atoms as required. Before the simulation, the system was relaxed and equilibrated. In the first equilibration, a constant number of particles, volume, and temperature (NVT) was chosen, and in the second equilibration, the number of particles, pressure, and temperature (NPT) was maintained constant. Equilibrations were done for 100 ps at 300 K using the Berendsen thermostat temperature coupling [59]. Then, the production of the trajectory of the molecular dynamic simulation was performed for 200 ns; the pressure coupling was set to 1 Barr and the thermostat coupling to 300 K [60].

2.9. Free Energy Calculations

The Molecular Mechanics Poisson–Boltzmann Surface Area (MM-PBSA) method was used to calculate the binding free energy of each candidate molecule to the FFA1 orthostatic binding pocket. The binding free energy was estimated using the g_mmpbsa tool [61] and Equation (1),

$$G_X = E_{MM} + G_{polar} + G_{nonpolar}, \tag{1}$$

where X is the ligand, receptor, or complex; E_{MM} is the vacuum molecular mechanics potential energy obtained for bonded and non-bonded interactions estimated using MM force field parameters; G_{polar} is the polar solvation energy calculated solving the Poisson–Boltzmann equation; $G_{nonpolar}$ is the non-polar solvation energy obtained using the solvent-accessible surface area (SASA) model. These energies were obtained for the protein, ligand, and complex. The trajectory between 10 and 200 ns was used, taking snapshots every 5 ns.

3. Results and Discussion

3.1. Dataset and Variable Selection

Descriptors extracted from the 3D structures of the 93 compounds reported by Christiansen et al. [17–20] were used to generate predictive models of agonist activity; the descriptors of the best model were used to split the data into training and test sets. The pEC$_{50}$ values cover more than 3 log activity distributions (from 4.79 to 8.04), which facilitates identification of descriptors that correlate with high agonist activity. The predictive models were obtained and evaluated with IBK, MLR, and Random Forest regression techniques based on the values of statistical parameters R^2 and Q^2_{CV} (see Table S2). Models 1 and 2 (M1 and M2), obtained with MLR, have the highest R^2 and Q^2_{CV}. Therefore, M1 (with 10 descriptors, Table S3) and M2 (with 11 descriptors, Table S3) were chosen for further analysis. A rational division of the molecules into training and test set was performed by applying the Ward's method's cluster analysis. The molecules were grouped into 10 clusters and approximately 25% were selected as a test set (Figure 1).

3.2. Applicability Domain

To detect outliers on the test set and provide reliable and accurate predictions, the applicability domain analysis was used. The applicability domain analysis used in this study is based on a consensus score of four methods. The score represents the fraction of the four methods that indicate a molecule is considered inside the applicability domain, with scores ranging from zero (molecule is identified as an outlier by all four methods) to one (molecule is inside the AD for all four methods) [21,62]. A score greater than 0.25 was used as the criteria for identifying compounds inside the AD. In M1, compounds 46 and 67 were identified as outside the AD and were removed from further analysis (Figure S1a). In contrast, all molecules in M2 met the criteria (Figure S1b).
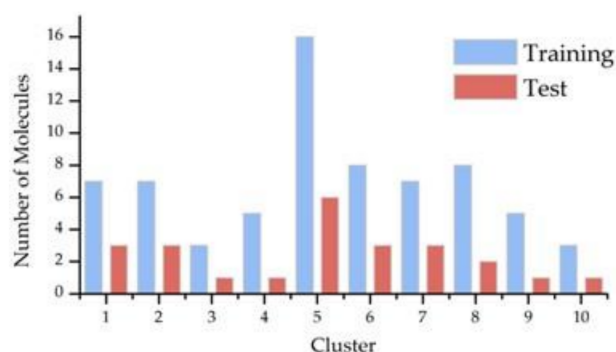
Figure 1. The dataset was divided in 10 clusters employing Ward's method to split into blue bars (training 74%) and red bars (test set 26%).

3.3. Model Validation

The outliers were removed from the test set to evaluate the performance of M1 and M2. The performance of M1 and M2 was evaluated using R^2, Q^2_{CV}, Q^2_{ext}, and mean absolute error (MAE) values. Values close to one for the statistic parameters R^2, Q^2_{CV}, and Q^2_{ext} indicate the models' goodness-of-fit and predictability (Figure 2a). In addition, a reliable prediction is considered if the value of MAE is smaller than 0.1 times the training set range in logarithm units (MAE $< 0.1 \times$ (pEC$_{50\text{-max}}$ − pEC$_{50\text{-min}}$)). The MAE values of both models are smaller than 0.3, meeting the criteria for a reliable prediction [63] (Figure 2b).
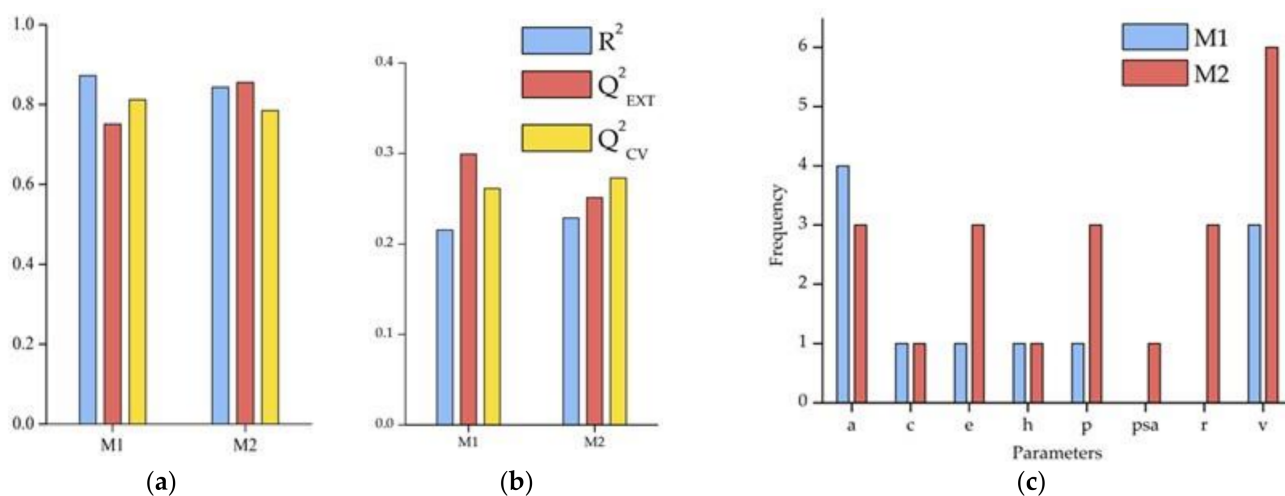

Figure 2. M1 and M2 obtained from SS_1213 and CSE_1393 set of descriptors. (**a**) presents the R^2, Q^2_{CV}, and Q^2_{EXT} values; (**b**) MAE values; and (**c**) is the frequency of the physicochemical properties within the descriptors for M1 and M2.

The descriptors were calculated using physicochemical properties, which are largely used to understand the nature of the compound for drug development. The descriptors of M1 and M2 contain the following physicochemical properties: *AlogP* (a), Charge (c), Electronegativity (e), Hardness (h), Polarizability (p), topological polar surface (psa), Refractivity (r), and Van der Waals volume (v). Figure 2c shows the number of times that property appears in M1 and M2 descriptors. *AlogP*, present in both models, quantifies molecular lipophilicity, which is crucial for developing FFA1 agonists [17–20]. On the other hand, the frequencies for the physicochemical properties "e", "p", and "v" are substantially higher in M2 than in M1.

Additional statistical parameters demonstrate proper fitting of the models. Models M1 and M2 are presented in Equations (2) and (3), respectively. Q^2_{boot} is the pEC$_{50}$ prediction obtained by a based perturbation. The values of $Q^2_{boot} > 0.7$ suggest that the dataset's

correlation fitting does not affect the perturbation. $a(R^2)$ and $a(Q^2)$ were obtained from Y-scrambling analysis and their small values suggest the pEC_{50} prediction does not occur by chance. M1 and M2 are well-fitted models based on the high value of the Fisher's statistical test ($F > 27$), the slight standard deviation ($s < 0.334$), and the correlation between predicted and experimental pEC_{50} (Figure 3).

$$pEC_{50\text{-M1}} = 0.518\,X_{3D1} - 0.132\,X_{3D2} - 0.325\,X_{3D3} + 0.002\,X_{3D4} + 0.038\,X_{3D5} - 0.024\,X_{3D6} + 162.563\,X_{3D7}$$
$$+ 7.409\,X_{3D8} + 0.045\,X_{3D9} + 0.222\,X_{3D10} + 3.3864 \tag{2}$$

$$R^2 = 0.872,\ Q^2_{\text{boot}} = 0.792,\ F = 39.47,\ s = 0.298,\ a(R^2) = 0.112,\ \text{and}\ a(Q^2) = -0.269$$

$$pEC_{50\text{-M2}} = -0.858\,Y_{3D1} + 0.056\,Y_{3D2} + 0.090\,Y_{3D3} + 31.113\,Y_{3D4} + 0.175\,Y_{3D5} - 1.512\,Y_{3D6} - 0.009\,Y_{3D7}$$
$$+ 0.620\,Y_{3D8} - 4.26\,Y_{3D9} - 0.069\,Y_{3D1O} - 0.059\,Y_{3D11} - 32.983 \tag{3}$$

$$R^2 = 0.843,\ Q^2_{\text{boot}} = 0.740,\ F = 27.74,\ s = 0.334,\ a(R^2) = 0.130,\ \text{and}\ a(Q^2) = -0.302$$
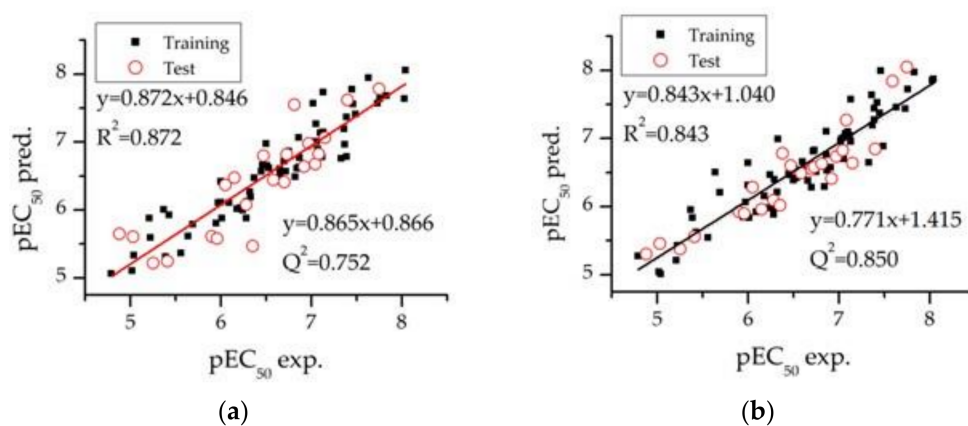

Figure 3. pEC_{50} calculated with M1 (**a**) and M2 (**b**) versus pEC_{50} experimental.

To analyze the collinearity of the descriptors, Pearson's coefficients were calculated for pairwise comparisons of the descriptors in M1 and M2 and are available in supporting information (Tables S4 and S5). The Pearson's coefficients for M1 (from -0.501 to 0.408) and M2 (from -0.330 to 0.453) suggest non-collinearity between the models' descriptors.

In addition, Tropsha's test (available at www.oecd.org (accessed on 18 February 2021)) validates the accuracy of the prediction of pEC_{50} values for both models. M1 and M2 met all the test requirements, suggesting an accurate prediction (Table 2). The model that will be considered for further analysis is based on the coverage and predictability of the external test set. Therefore, based on the 100% coverage on the AD, M2 will be considered to predict the pEC_{50} for the screening databases.

3.4. Screening of the DrugBank 5.1.7 and DiaNat Databases

With the aim to find new drug candidates for treating T2DM, molecules from relevant databases DiaNat [27] and DrugBank 5.1.7 [28] were screened using M2. A total of 31 compounds from DiaNat-DB met our criteria for the applicability domain with pEC_{50} values estimated to be from 5.5 to 8.0. pEC_{50} values and the SMILES are available in Table S6. The molecules with highest predicted affinity for FFA1 ($pEC_{50} > 7.5$) are nuciferine, γ-mangostin, morolic acid, curcumin, and 3-*O*-acetyloleanolic acid.

Table 2. Validation based on the Tropsha's test for M1 and M2. Val is the abbreviation of validation.

M1	Cross-Validation		External Validation	
Criterion	**Result**	**Assessment**	**Result**	**Assessment**
$R^2 > 0.6$	0.872	PASS	0.872	PASS
$Q^2_{Val} > 0.5$	0.816	PASS	0.752	PASS
$(Q^2_{Val} - R_0^2)/Q^2_{Val} < 0.1$	0.001	PASS	0.023	PASS
$(Q^2_{Val} - R_0'^2)/Q^2_{Val} < 0.1$	0.037	PASS	0.024	PASS
$abs(R_0^2 - R_0'^2) < 0.1$	0.030	PASS	0.001	PASS
$0.85 < k < 1.15$	0.999	PASS	0.998	PASS
$0.85 < k' < 1.15$	0.998	PASS	0.998	PASS
M2	**Cross-Validation**		**External Validation**	
Criterion	**Result**	**Assessment**	**Result**	**Assessment**
$R^2 > 0.6$	0.843	PASS	0.843	PASS
$Q^2_{Val} > 0.5$	0.778	PASS	0.850	PASS
$(Q^2_{Val} - R_0^2)/Q^2_{Val} < 0.1$	0.002	PASS	0.037	PASS
$(Q^2_{Val} - R_0'^2)/Q^2_{Val} < 0.1$	0.054	PASS	0.001	PASS
$abs(R_0^2 - R_0'^2) < 0.1$	0.041	PASS	0.032	PASS
$0.85 < k < 1.15$	0.998	PASS	1.005	PASS
$0.85 < k' < 1.15$	0.999	PASS	0.993	PASS

A total of 282 compounds from DrugBank 5.1.7 met our criteria for the applicability domain. Their predicted pEC_{50} values are from 3.77 to 9.35, and SMILES are available in Table S7. It is important to note that 40 compounds of the DrugBank database have pEC_{50} values greater than 8, with the antipsychotic Haloperidol (DB00502), Vitamin K1 (DB01022), and carotenoid Zeaxanthin (DB11176) having the most significant values.

Further results about ADME predictions, molecular docking, molecular dynamics simulations of the most suitable FFA1 agonist candidates of the dataset (training/test set), and screening (DiaNat, and DrugBank) databases, are presented in the following sections.

3.5. Absorption, Distribution, Metabolism, and Excretion (ADME) Predictions

SwissADME (http://www.swissadme.ch (accessed on 18 February 2021)) was used to calculate physicochemical properties, lipophilicity, water-solubility, pharmacokinetics, drug-likeness, medicinal chemistry properties, and bioavailability score of the molecules. Bioavailability is a fast screening in the compounds to evaluate the possibility to be considered an oral drug [64]. All the compounds have a high bioavailability score of 0.85 in the dataset. The lipophilicity obtained from Wildman and Crippen method (WLOGP) and the topological polar surface area (TPSA) were used to construct the BOILED-Egg (Figure 4). The Boiled egg is a plot of WLOGP vs. TPSA and predicts the gastrointestinal absorption (HIA) and brain penetration (BBB). The four molecules (46, 50, 51, and 53) in the grey region are predicted to have poor intestinal absorption and brain penetration. P-glycoprotein substrates (Pg-p) is a restrictive barrier to maintain homeostasis of the brain and is key among ATP-binding cassette transporters. The red and blue dots are p-glycoprotein substrates (PGP+) and *P*-glycoprotein non-substrates (PGP−), respectively [65].

The ADME properties of the eight compounds in the dataset with the highest biological activity ($pEC_{50} > 7.4$), BBB permeation, and high HIA, are shown in Table 3 along with solubility and lipophilicity properties. The searching of possible FFA1 agonists with low lipophilicity has been highlighted by Christiansen et al., who focused on derivatives of TUG-424 (Molecule 15) [20]. Lipophilicity and water solubility have been shown to facilitate drug formulation and handling [66]. In this regard, Molecules considered for further analysis must have a log $P_{o/w}$ value smaller than 3.83. Molecules 15, 52, 49, 48, and 47 are discarded for further investigation to be too lipophilic (Consensus Log $P_{o/w} > 3.83$). Water solubility was evaluated by three models, the first was ESOL [67]; the second was adapted from Ali et al. [68]; the third was developed by SILICOS-IT [66]. The qualitative

scale to estimate water solubility is based on the log S scale: highly soluble > 0 > very soluble > −2 > soluble > −4 > moderately soluble > −6 > poorly soluble > −10 > insoluble. Compounds 91 (test set), 92 (training set), and 93 (training set) are either moderately soluble or soluble in at least two solubility methods. These compounds have smaller lipophilicity attributed to the incorporation of polar methoxy (compound 92) or cyanide (compounds 91 and 93) groups.
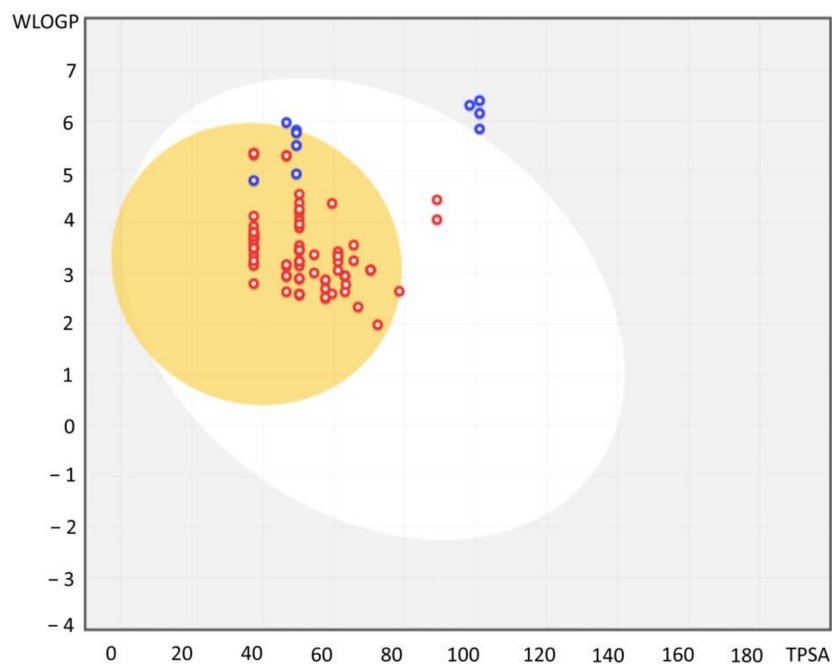

Figure 4. BOILED-EGG plot showing the probability of human intestinal absorption (HIA) and blood–brain barrier (BBB) permeation of the studied molecules.

Table 3. Properties calculated from ADME, solubility (S = Soluble, Ms = Moderately soluble, Ps = Poorly soluble), and log $P_{o/w}$ consensus.

Molecule	pEC_{50}	ESOL Class	Ali Class	Silicos-IT Class	Consensus Log $P_{o/w}$
52	8.03	Ms	Ps	Ps	4.93
49	7.75	Ms	Ps	Ps	5.24
48	7.73	Ms	Ms	Ps	4.59
47	7.63	Ms	Ps	Ps	5.22
15	7.49	Ms	Ms	Ms	3.83
93	7.45	S	Ms	Ms	3.37
92	7.42	Ms	Ms	Ms	3.51
91	7.4	Ms	Ms	Ms	3.57

Regarding the compounds from the screening database, approximately 90.5% of the screened molecules are BBB permeable and have high HIA. Same as mentioned above for the dataset, the molecules in the screening to be considered for further analysis must meet high activity ($pEC_{50} \geq 7.4$) and low lipophilicity (Consensus log $P_{o/w} \leq 3.83$). Consequently, 25 compounds from the Drugbank and 1 compound from the DiaNat database met these requirements and were considered for further analysis.

3.6. Molecular Docking

Molecular docking was used to characterize the binding mode and steric fit of the ligands into the orthosteric site of FFA1, as well as to gain insights into protein–ligand interactions involved. Negative docking scores of the database compounds, ranging from −8.1 to −11.0 kcal/mol (Table S1), suggest a favorable fit into the active site. The docking

scores of the most potent compounds, i.e., 91 (−9.4 kcal/mol), 92 (−9.2 kcal/mol), and 93 (−10.3 kcal/mol), are comparable to the positive controls TAK-875 (−9.9 kcal/mol) and 15 (−10.2 kcal/mol).

The experimental configuration of TAK-875 bound to FFA1 (PDB: 4PHU) and the predicted docked configuration obtained in this work for the same compound were compared (Figure 5a). The molecular structure of the experimental structure is mostly superimposed with the docked configuration (RMSD = 1.90 Å), except at the methylsulfonyl tail of the molecule after the ether group that differs substantially. However, this is unlikely to be of consequence since the tail is very flexible and lies outside the orthosteric pocket, so it would not affect the activation occurring inside the active site. On the other end of the molecule, it is observed that the carboxylic acid of the docked structure presents a small shift compared to the experimental one.

(a)   **(b)**
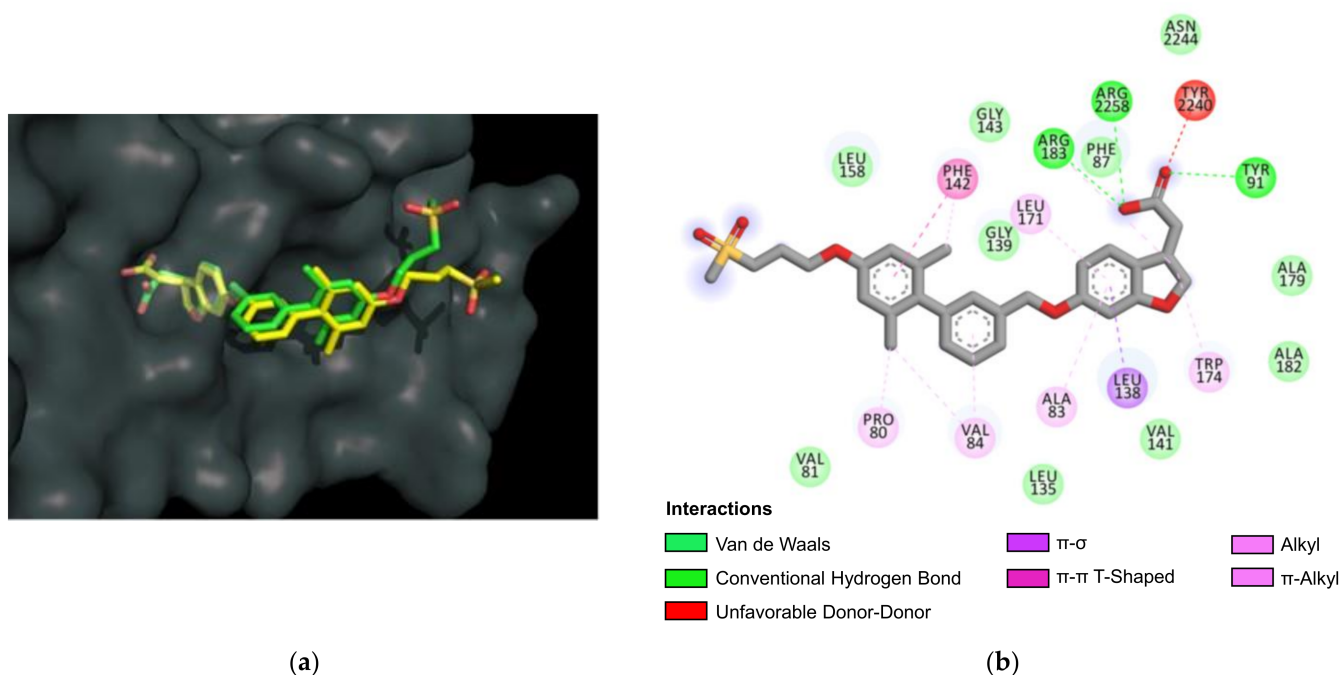
Figure 5. (**a**) Experimental (yellow) and docked (green) configurations of TAK-875 bound to FFA1. (**b**) 2D graphic of the interaction between TAK-875 and FFA1.

The interactions of the positive controls occur mainly with Ala-83, Val-84, Tyr-91, Leu-138, Phe-142, Leu-171, and Arg-183. The interactions of TAK-875 and compound 15 with FFA1 are shown in Figure 5b and Figure S2a, respectively. The carboxylic group in TAK-875 forms hydrogen bonds (HBs) with Tyr-91, Arg-183, and Arg-258. Arg183 and Arg-258 were reported in the literature as essential to anchor ligands containing carboxylate groups, while Tyr91 was reported to produce aromatic/hydrophobic interactions with the ligand [26].

Molecular docking was also performed on the 26 compounds resulting from the screening. The docking scores are presented in Table S8. Seventeen compounds were taken out from the study because they stay outside the active site. Furthermore, suprofen and carbocromen were also removed due to their severe side effect leading to their market withdrawal. The docking scores of the remaining seven compounds (−8.5 to −10.3 kcal/mol) suggest favorable FFA1-ligand interactions (Figure S3). These compounds are anileridine (DB00913), a synthetic opioid; bromfenac (DB00963), a non-steroidal anti-inflammatory; bilastine (DB11591), an antihistamine drug; sulfinpyrazone (DB01138), an uricosuric; fenofibric acid (DB13873), a lipid-lowering agent conditions including hypertriglyceridemia, mixed dyslipidemia, and primary hyperlipidemia [69]; indacaterol (DB05039) used in situations such as asthma and chronic obstructive pulmonary diseases [70]; and

curcumin (DBDB11672), an antioxidant, hypoglycemic, wound-healing, and antimicrobial compound [71].

The interactions between ligands and amino acids linked to FFA1 activation, high potency, and stabilization are shown in Figure 6. HBs are essential in the stabilization of the ligand–protein complex and FFA1 activation. Tyr-91, Arg-183, and Arg-258 were reported as relevant residues for the activation of the FFA1 [26]. Tyr-91 forms HB with compound 15, TAK-875, bilastine, and bromfenac. Arg-183 forms HBs with all compounds from the dataset and fenofibric acid. Arg-258 form HBs with compound 91, 92, TAK-875, and fenofibric acid. Interestingly, compounds 91, 92, and 93, the most actives compounds in the dataset, do not form hydrogen bonds with Tyr-91. On the other hand, Val-84 and Leu-138 seem to play an essential role during binding as they form π interactions with most of the compounds shown in Figure 6.
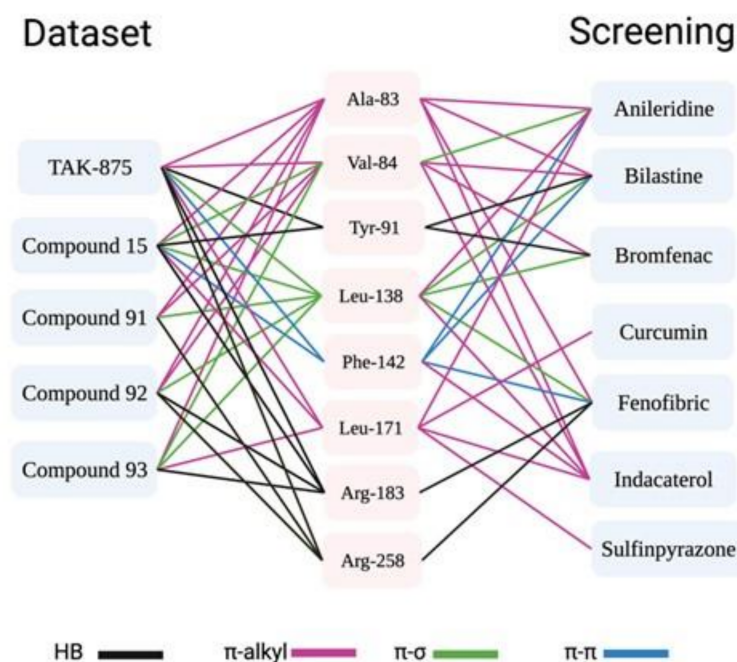
Figure 6. Interactions between FFA1 and selected molecules of the database and screening (created with BioRender.com).

3.7. Molecular Dynamic Simulations

Molecular dynamics simulations were run for 200 ns to explore the trajectory of the binding process of five compounds from the dataset (15, 91, 92, 93, and TAK-875) and the seven compounds selected from the screening. To determine the reliability of the calculation along the trajectory, the changes in the potential energy, density, pressure, volume, temperature, and box size plots were analyzed.

The root-mean-square deviation (RMSD) was analyzed to explore the stability of the FFA1-ligand complex during the trajectory. Both the RMSD of the FFA1 (Figure 7) and the ligands (Figure 8) are expected to change during the simulation as the protein and the ligand reach a stable complex conformation. When the system reaches a stable configuration, fluctuations in the RMSD should decrease, meaning equilibrium is achieved. FFA1 reaches equilibrium in most of the FFA1-ligand complexes after 60 ns. The FFA1-TAK-875 complex was taken from experimental data and started from the equilibrium. The lowest RMSD values in the FFA1-TAK-875 complex validate the feasibility of the simulations.
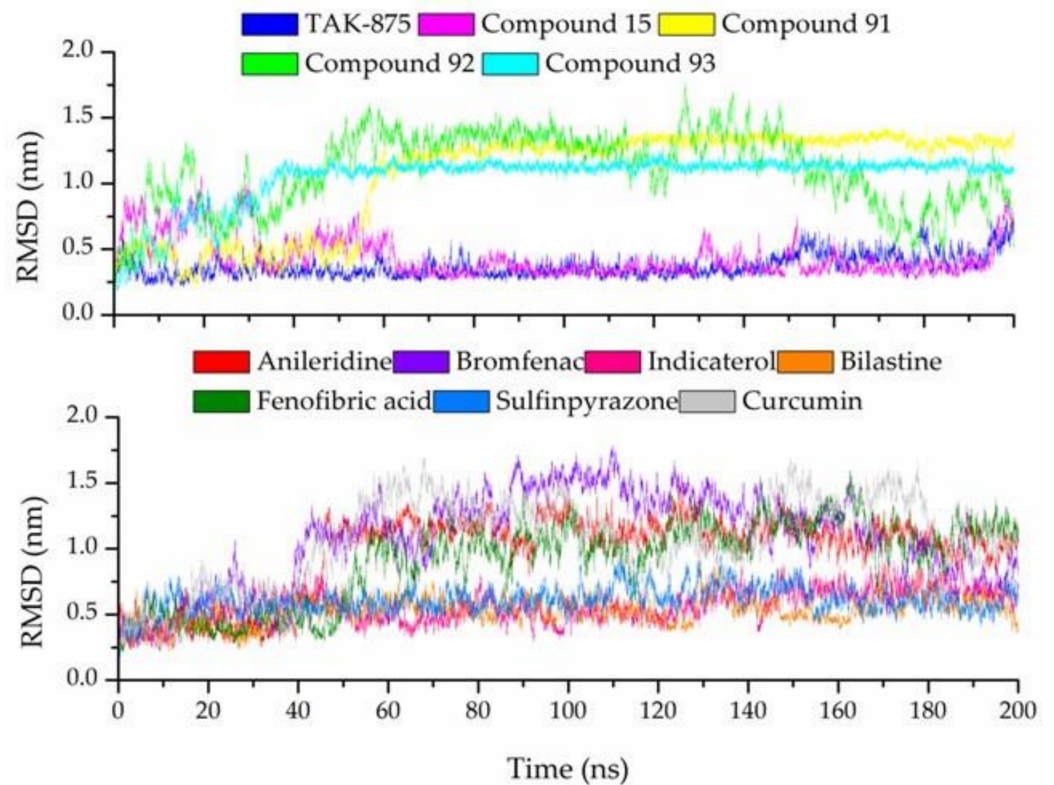
Figure 7. RMSD of (**top**) FFA1 in complex with the ligands of the dataset (**top**) and of the screening (**bottom**) during the 200 ns of MD simulation.
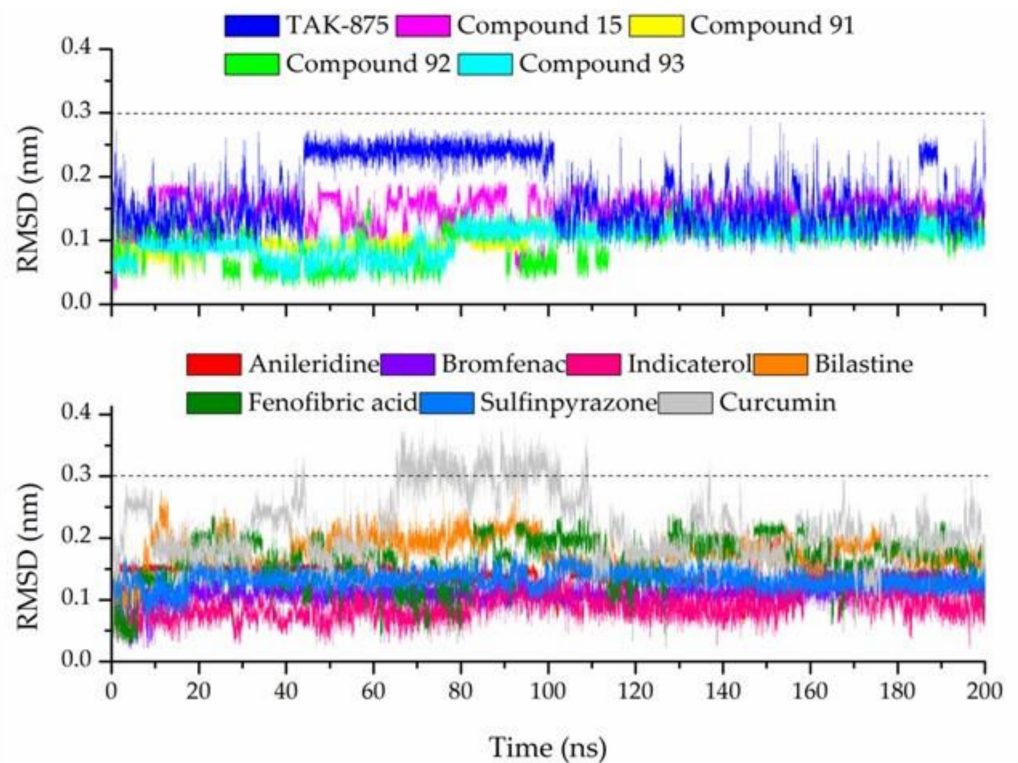

Figure 8. RMSD of the ligands of the dataset (**top**) and the screening (**bottom**) during the 200 ns MD simulation.

The RMSD values of the ligands below the threshold (dotted line 0.30 nm) mean the optimized structures found on the docking analysis were appropriate (Figure 8). The RMSD of the ligands suggests the stability of the system during the trajectory. TAK-875 RMSD "jumps" between 50–100 ns and 185–190 ns due to a fluctuation on the methylsulfonyl tail. Curcumin RMSD goes above the threshold moving outside the active site during the simulation. Therefore, curcumin was taken out for further analysis.

Potent agonists showed interactions with amino acids that might be essential for the activation of FFA1. The number of hydrogen bonds (HBs) in the FFA1-ligand complex was analyzed along with the occupancy (number of times that a HB appears during the simulation) with residues Tyr-91, Arg-183, Asn-244, and Arg-258. Most of the compounds form between 1 to 6 HBs (Figure 9). Sulfinpyrazone comprises the highest number of HBs (24), and compound 92 forms the least. The occupancies confirm similar HB formation to the molecular docking study. Most of the compounds form interactions with Tyr-91, which is related to the activation of FFA1 and higher potency of the ligand [26]. However, compound 93, one of the most potent compounds tested in vitro, did not interact with Tyr-91. This compound has the highest occupancy with Arg-183 (63.0%). Asn-244 was reported to be important in anchoring the carboxylic acid [26]. Asn-244 did not appear in the docking study but presents important occupancies with compounds 15 (40%) and 93 (17%). Therefore, Arg-183 and Asn-244 may have an essential role in FFA1 activation.
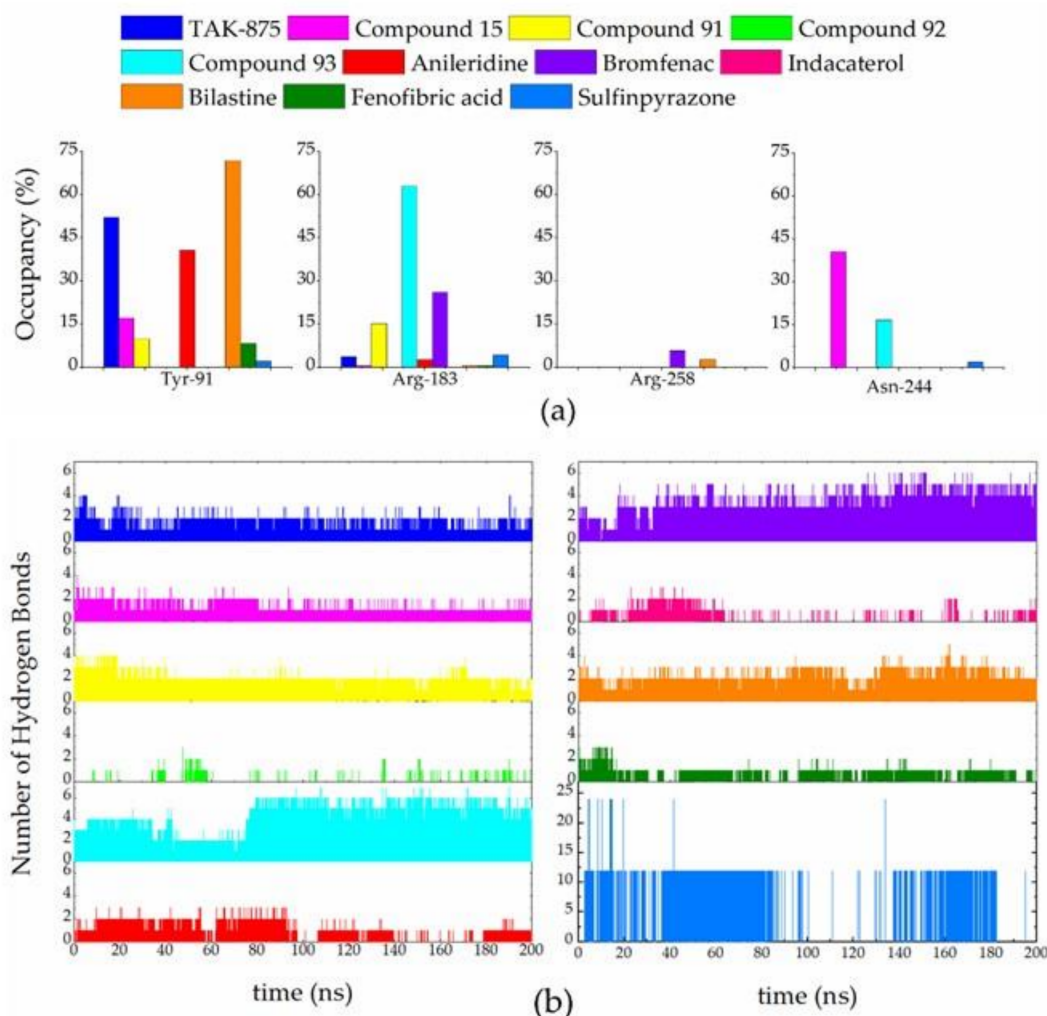

Figure 9. (**a**) Occupancies and (**b**) number of hydrogen bonds between FFA1 and the studied compounds during the 200 ns simulation.

Lennard-Jones energy and short-range Coulomb potential quantify the interaction energy of the protein–ligand complex (Figures S4 and S5). The short-range Lennard-Jones energy profile for all the models (-55 to -35 kcal/mol) indicates a similar interaction between the ligands and FFA1. Bilastine presented a -45 kcal/mol value during the first half and a value of -60 kcal/mol in the second half of the simulation, which suggests a more stable interaction was achieved in the second half. The ligands present a short-range Coulomb potential in accordance with the HBs formation during the dynamics. In this sense, compound 92 and indacaterol have the highest values, while compound 93 and bromfenac are the lowest. According to short-range Lennard-Jones energy and short-range Coulomb potential, all the compounds managed to interact favorably with FFA1, being possible agonist candidates.

The ligand–protein interaction was analyzed applying the MMPBSA approach [61], in which it is possible to estimate the free binding energy based on van der Waals, electrostatic, and solvent accessible surface area (SASA) energies (Table 4). The negative values found for the four energies agree that there is a favorable interaction between FFA1 and the ligands. Similar binding energy values were obtained for all the datasets' systems, being TAK-875, the compound with the highest affinity (-30.88 kcal/mol). For the ligands of the screening, the range of the binding energy goes from -15 kcal/mol in bromfenac to -37 kcal/mol in bilastine. Although bromfenac presents the lowest binding energy value, its predicted pEC_{50} is the highest, comparing to TAK-875 (8.47 vs. 8.45).

Table 4. Van der Waals, electrostatic, SASA, and binding energy (BE) (in kcal/mol) of the studied compounds.

Compound	Van der Waals Energy	Electrostatic Energy	SASA Energy	Binding Energy	pEC_{50} Pred.
TAK-875	-55.80	-9.78	-5.28	-30.88	8.45
15	-38.75	-5.40	-3.97	-25.15	6.88
91	-42.97	-8.39	-4.07	-26.96	6.84
92	-40.11	-2.14	-4.00	-28.70	7.53
93	-43.84	-18.10	-4.29	-28.61	7.37
Anileridine	-50.91	-7.76	-4.92	-32.69	7.82
Bromfenac	-38.08	-32.20	-3.80	-15.22	8.47
Sulfinpyrazone	-53.32	-5.75	-5.14	-24.49	7.55
Indacaterol	-43.90	-0.91	-4.43	-28.63	7.41
Bilastine	-59.13	-16.58	-5.83	-36.97	7.57
Fenofibric acid	-35.25	-4.67	-3.70	-20.70	7.92

The structures of the six compounds were compared to the TAK-875 structure to spot similarities and differences that could be essential for interaction at the active site (Figure 10). Interestingly, the carboxylic group in TAK-875 is also present in bromfenac, bilastine, and fenofibric acid. In anileridine, if a carboxyl group replaces the ester group, it may produce HBs, enhancing FFA1 activation. The TAK-875 benzene ring in the dihydrobenzofuran group interacts with Ala-83, Leu-138, and Leu-171 (Figure 5b). The benzene ring near the carboxyl group in bromfenac, bilastine, and fenofibric acid may produce the same interactions with those amino acids. Furthermore, anileridine, sulfinpyrazone, and indacaterol also present cyclic motifs in their structure that interact with Ala-83, Leu-138, and Leu-171. Finally, Val-84 that interacts with the second benzene ring in TAK-875 can interact with any cyclic structures on the left side of the six studies compounds.
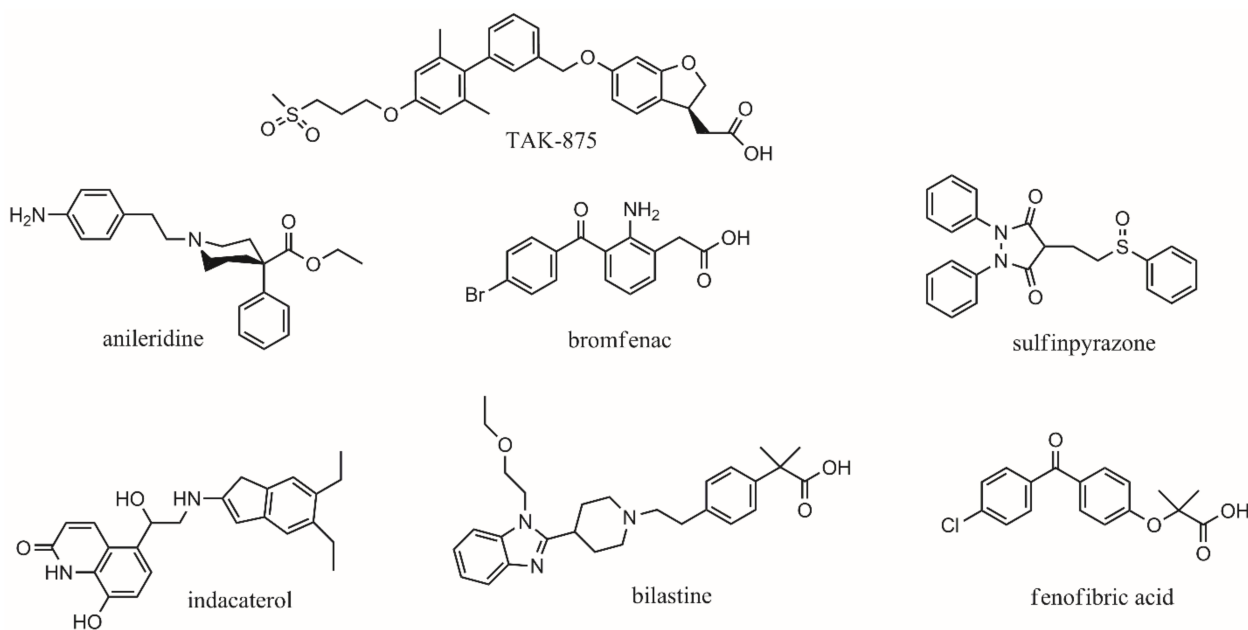
Figure 10. 2D structure of TAK-875 and the best compounds resulted from the screening.

Based on these results, bilastine presents the best characteristics, including the lowest binding energy (-36.97 kcal/mol), good lipophilicity (3.76), and a strong HB with Tyr-91 (71.9% occupancy) to be considered for in vitro analysis as an FFA1 agonist or as a lead compound to develop next-generation drugs against T2DM. Bromfenac is also an excellent option as a lead molecule due to its highest predicted pEC50 value. Bilastine and bromfenac are drugs of interest to treat diabetes-related diseases [72,73]. In vivo studies suggested bilastine as therapy for diabetic nephropathy, a common microvascular complication in patients diagnosed with diabetes mellitus [72]. Clinical trials reported the long-term benefits of bromfenac treating diabetic macular edema, a cause of loss of vision in patients with diabetes [73]. Finally, fenofibric acid is an interesting lead as changes may be introduced to produce tighter interactions with the FFA1 active site, developing a dual-target drug to treat T2DM and hypertriglyceridemia.

4. Conclusions

A dataset containing 93 compounds was split into training and test set and modelled employing different machine-learning techniques. M1 and M2 models, evaluated with MLR, showed the best statistical parameters: $R^2 = 0.872$, $Q^2_{CV} = 0.812$, and $Q^2_{ext} = 0.751$, for M1; and $R^2 = 0.843$, $Q^2_{CV} = 0.875$, and $Q^2_{ext} = 0.850$, for M2. Both models comply with the Tropsha's test validation. M2 present the best coverage in the applicability domain analysis for the test set (100%). Therefore, it was employed for the screening of two relevant datasets (DiaNat and DrugBank). Based on a high activity ($pEC_{50} \geq 7.4$) and low lipophilicity (Consensus log $P_{o/w} \leq 3.83$) criterion, 26 compounds were selected as promising drugs. An exhaustive analysis on the protein–ligand interaction was done in the training, test set, and screened compounds using molecular docking and molecular dynamics tools. It was found that the interactions with Tyr-91, Leu-138, Leu-171, Arg-183, Ala-83, Val-84, Asn-244, and Arg-258 are crucial for FFA1 activation. The structure of the possible candidates presents different cores, which can be further explored to find new FFA1 agonists. These candidates are FDA-approved drugs, and future studies can explore their biological activity as FFA1 agonists.

References

1. Huang, Y.; Karuranga, S.; Malanda, B.; Williams, D.R.R. Call for data contribution to the IDF Diabetes Atlas 9th Edition 2019. *Diabetes Res. Clin. Pract.* **2018**, *140*, 351–352. [CrossRef] [PubMed]
2. Eizirik, D.L.; Pasquali, L.; Cnop, M. Pancreatic beta-cells in type 1 and type 2 diabetes mellitus: Different pathways to failure. *Nat. Rev. Endocrinol.* **2020**, *16*, 349–362. [CrossRef] [PubMed]
3. Chatterjee, S.; Khunti, K.; Davies, M.J. Type 2 diabetes. *Lancet* **2017**, *389*, 2239–2251. [CrossRef]
4. DeFronzo, R.A.; Ferrannini, E.; Groop, L.; Henry, R.R.; Herman, W.H.; Holst, J.J.; Hu, F.B.; Kahn, C.R.; Raz, I.; Shulman, G.I.; et al. Type 2 diabetes mellitus. *Nat. Rev. Dis. Primers* **2015**, *1*, 15019. [CrossRef] [PubMed]
5. Turner, R.C.; Cull, C.A.; Frighi, V.; Holman, R.R. Glycemic control with diet, sulfonylurea, metformin, or insulin in patients with type 2 diabetes mellitus: Progressive requirement for multiple therapies (UKPDS 49). UK Prospective Diabetes Study (UKPDS) Group. *JAMA* **1999**, *281*, 2005–2012. [CrossRef] [PubMed]
6. Sharma, N.; Bhagat, S.; Chundawat, T.S. Recent Advances in Development of GPR40 Modulators (FFA1/FFAR1): An Emerging Target for Type 2 Diabetes. *Mini Rev. Med. Chem.* **2017**, *17*, 947–958. [CrossRef]
7. Houthuijzen, J.M. For Better or Worse: FFAR1 and FFAR4 Signaling in Cancer and Diabetes. *Mol. Pharmacol.* **2016**, *90*, 738–743. [CrossRef]
8. Teng, D.; Chen, J.; Li, D.; Wu, Z.; Li, W.; Tang, Y.; Liu, G. Computational Insights into Molecular Activation and Positive Cooperative Mechanisms of FFAR1 Modulators. *J. Chem. Inf. Model.* **2020**, *60*, 3214–3230. [CrossRef] [PubMed]
9. Choi, Y.J.; Shin, D.; Lee, J.Y. G-protein coupled receptor 40 agonists as novel therapeutics for type 2 diabetes. *Arch. Pharm. Res.* **2014**, *37*, 435–439. [CrossRef]
10. Kaku, K.; Enya, K.; Nakaya, R.; Ohira, T.; Matsuno, R. Long-term safety and efficacy of fasiglifam (TAK-875), a G-protein-coupled receptor 40 agonist, as monotherapy and combination therapy in Japanese patients with type 2 diabetes: A 52-week open-label phase III study. *Diabetes Obes. Metab.* **2016**, *18*, 925–929. [CrossRef] [PubMed]

11. Marcinak, J.F.; Munsaka, M.S.; Watkins, P.B.; Ohira, T.; Smith, N. Correction to: Liver Safety of Fasiglifam (TAK-875) in Patients with Type 2 Diabetes: Review of the Global Clinical Trial Experience. *Drug Saf.* **2018**, *41*, 1431–1437. [CrossRef]

12. Edfalk, S.; Steneberg, P.; Edlund, H. Gpr40 is expressed in enteroendocrine cells and mediates free fatty acid stimulation of incretin secretion. *Diabetes* **2008**, *57*, 2280–2287. [CrossRef]

13. Kotarsky, K.; Nilsson, N.E.; Flodgren, E.; Owman, C.; Olde, B. A human cell surface receptor activated by free fatty acids and thiazolidinedione drugs. *Biochem. Biophys. Res. Commun.* **2003**, *301*, 406–410. [CrossRef]

14. Stoddart, L.A.; Smith, N.J.; Milligan, G. International Union of Pharmacology. LXXI. Free fatty acid receptors FFA1, -2, and -3: Pharmacology and pathophysiological functions. *Pharm. Rev.* **2008**, *60*, 405–417. [CrossRef] [PubMed]

15. Brown, S.P.; Dransfield, P.J.; Vimolratana, M.; Jiao, X.; Zhu, L.; Pattaropong, V.; Sun, Y.; Liu, J.; Luo, J.; Zhang, J.; et al. Discovery of AM-1638: A Potent and Orally Bioavailable GPR40/FFA1 Full Agonist. *ACS Med. Chem. Lett.* **2012**, *3*, 726–730. [CrossRef]

16. Holliday, N.; Watson, S.-J.; Brown, A. Drug Discovery Opportunities and Challenges at G Protein Coupled Receptors for Long Chain Free Fatty Acids. *Front. Endocrinol.* **2012**, *2*, 112. [CrossRef]

17. Christiansen, E.; Urban, C.; Merten, N.; Liebscher, K.; Karlsen, K.K.; Hamacher, A.; Spinrath, A.; Bond, A.D.; Drewke, C.; Ullrich, S. Discovery of potent and selective agonists for the free fatty acid receptor 1 (FFA1/GPR40), a potential target for the treatment of type II diabetes. *J. Med. Chem.* **2008**, *51*, 7061–7064. [CrossRef] [PubMed]

18. Christiansen, E.; Urban, C.; Grundmann, M.; Due-Hansen, M.E.; Hagesaether, E.; Schmidt, J.; Pardo, L.; Ullrich, S.; Kostenis, E.; Kassack, M. Identification of a potent and selective free fatty acid receptor 1 (FFA1/GPR40) agonist with favorable physicochemical and in vitro ADME properties. *J. Med. Chem.* **2011**, *54*, 6691–6703. [CrossRef]

19. Christiansen, E.; Due-Hansen, M.E.; Urban, C.; Grundmann, M.; Schröder, R.; Hudson, B.D.; Milligan, G.; Cawthorne, M.A.; Kostenis, E.; Kassack, M.U. Free fatty acid receptor 1 (FFA1/GPR40) agonists: Mesylpropoxy appendage lowers lipophilicity and improves ADME properties. *J. Med. Chem.* **2012**, *55*, 6624–6628. [CrossRef]

20. Christiansen, E.; Due-Hansen, M.E.; Urban, C.; Grundmann, M.; Schmidt, J.; Hansen, S.V.; Hudson, B.D.; Zaibi, M.; Markussen, S.B.; Hagesaether, E. Discovery of a potent and selective free fatty acid receptor 1 agonist with low lipophilicity and high oral bioavailability. *J. Med. Chem.* **2013**, *56*, 982–992. [CrossRef]

21. Mora, J.R.; Marrero-Ponce, Y.; García-Jacas, C.R.; Suarez Causado, A. Ensemble models based on QuBiLS-MAS features and shallow learning for the prediction of drug-induced liver toxicity: Improving deep learning and traditional approaches. *Chem. Res. Toxicol.* **2020**. [CrossRef]

22. Cabrera, N.; Mora, J.R.; Márquez, E.; Flores-Morales, V.; Calle, L.; Cortés, E. QSAR and molecular docking modelling of anti-leishmanial activities of organic selenium and tellurium compounds. *SAR QSAR Environ. Res.* **2020**, *32*, 29–50. [CrossRef] [PubMed]

23. Gaba, V.; Rani, K.; Gupta, M.K. QSAR study on 4-alkynyldihydrocinnamic acid analogs as free fatty acid receptor 1 agonists and antidiabetic agents: Rationales to improve activity. *Arab. J. Chem.* **2019**, *12*, 1758–1764. [CrossRef]

24. Srivastava, A.; Yano, J.; Hirozane, Y.; Kefala, G.; Gruswitz, F.; Snell, G.; Lane, W.; Ivetac, A.; Aertgeerts, K.; Nguyen, J. High-resolution structure of the human GPR40 receptor bound to allosteric agonist TAK-875. *Nature* **2014**, *513*, 124–127. [CrossRef]

25. Tikhonova, I.G.; Sum, C.S.; Neumann, S.; Thomas, C.J.; Raaka, B.M.; Costanzi, S.; Gershengorn, M.C. Bidirectional, iterative approach to the structural delineation of the functional "chemoprint" in GPR40 for agonist recognition. *J. Med. Chem.* **2007**, *50*, 2981–2989. [CrossRef]

26. Sum, C.S.; Tikhonova, I.G.; Neumann, S.; Engel, S.; Raaka, B.M.; Costanzi, S.; Gershengorn, M.C. Identification of residues important for agonist recognition and activation in GPR40. *J. Biol. Chem.* **2007**, *282*, 29248–29255. [CrossRef] [PubMed]

27. Madariaga-Mazón, A.; Naveja, J.J.; Medina-Franco, J.L.; Noriega-Colima, K.O.; Martinez-Mayorga, K. DiaNat-DB: A molecular database of antidiabetic compounds from medicinal plants. *RSC Adv.* **2021**, *11*, 5172–5178. [CrossRef]

28. Wishart, D.S.; Feunang, Y.D.; Guo, A.C.; Lo, E.J.; Marcu, A.; Grant, J.R.; Sajed, T.; Johnson, D.; Li, C.; Sayeeda, Z.; et al. DrugBank 5.0: A major update to the DrugBank database for 2018. *Nucleic Acids Res.* **2017**, *46*, D1074–D1082. [CrossRef]

29. Sliwoski, G.; Kothiwale, S.; Meiler, J.; Lowe, E.W., Jr. Computational methods in drug discovery. *Pharm. Rev.* **2013**, *66*, 334–395. [CrossRef]

30. Li, Z.; Hu, L.; Wang, X.; Zhou, Z.; Deng, L.; Xu, Y.; Zhang, L. Design, synthesis, and biological evaluation of novel dual FFA1 (GPR40)/PPARδ agonists as potential anti-diabetic agents. *Bioorg. Chem.* **2019**, *92*, 103254. [CrossRef]

31. Li, Z.; Pan, M.; Su, X.; Dai, Y.; Fu, M.; Cai, X.; Shi, W.; Huang, W.; Qian, H. Discovery of novel pyrrole-based scaffold as potent and orally bioavailable free fatty acid receptor 1 agonists for the treatment of type 2 diabetes. *Bioorg. Med. Chem.* **2016**, *24*, 1981–1987. [CrossRef]

32. Rappe, A.K.; Casewit, C.J.; Colwell, K.S.; Goddard, W.A.; Skiff, W.M. UFF, a full periodic table force field for molecular mechanics and molecular dynamics simulations. *J. Am. Chem. Soc.* **1992**, *114*, 10024–10035. [CrossRef]

33. Garcia-Jacas, C.R.; Marrero-Ponce, Y.; Acevedo-Martinez, L.; Barigye, S.J.; Valdes-Martini, J.R.; Contreras-Torres, E. QuBiLS-MIDAS: A parallel free-software for molecular descriptors computation based on multilinear algebraic maps. *J. Comput. Chem.* **2014**, *35*, 1395–1409. [CrossRef] [PubMed]

34. García-Jacas, C.R.; Marrero-Ponce, Y.; Vivas-Reyes, R.; Suárez-Lezcano, J.; Martinez-Rios, F.; Terán, J.E.; Aguilera-Mendoza, L. Distributed and multicore QuBiLS-MIDAS software v2.0: Computing chiral, fuzzy, weighted and truncated geometrical molecular descriptors based on tensor algebra. *J. Comput. Chem.* **2020**, *41*, 1209–1227. [CrossRef]

35. Hall, M.; Frank, E.; Holmes, G.; Pfahringer, B.; Reutemann, P.; Witten, I.H. The WEKA data mining software: An update. *ACM SIGKDD Explor. Newsl.* **2009**, *11*, 10–18. [CrossRef]

36. Murtagh, F.; Legendre, P. Ward's Hierarchical Agglomerative Clustering Method: Which Algorithms Implement Ward's Criterion? *J. Classif.* **2014**, *31*, 274–295. [CrossRef]

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

38. Edraki, N.; Das, U.; Hemateenejad, B.; Dimmock, J.R.; Miri, R. Comparative QSAR analysis of 3, 5-bis (arylidene)-4-piperidone derivatives: The development of predictive cytotoxicity models. *Iran. J. Pharm. Res.* **2016**, *15*, 425.

39. Cabrera, N.; Mora, J.R.; Marquez, E.A. Computational molecular modeling of Pin1 inhibition activity of quinazoline, benzophenone, and pyrimidine derivatives. *J. Chem.* **2019**, *2019*, 2954250. [CrossRef]

40. Khan, P.M.; Roy, K. Current approaches for choosing feature selection and learning algorithms in quantitative structure–activity relationships (QSAR). *Expert Opin. Drug Discov.* **2018**, *13*, 1075–1089. [CrossRef]

41. Hall, M.A. Correlation-Based Feature Selection for Machine Learning. Ph.D. Thesis, The University of Waikato, Hamilton, New Zealand, 1999.

42. Hanser, T.; Barber, C.; Marchaland, J.F.; Werner, S. Applicability domain: Towards a more formal definition. *SAR QSAR Environ. Res.* **2016**, *27*, 865–881. [CrossRef] [PubMed]

43. Roy, K.; Kar, S.; Ambure, P. On a simple approach for determining applicability domain of QSAR models. *Chemom. Intell. Lab. Syst.* **2015**, *145*, 22–29. [CrossRef]

44. Jaworska, J.; Nikolova-Jeliazkova, N. How can structural similarity analysis help in category formation? *SAR QSAR Environ. Res.* **2007**, *18*, 195–207. [CrossRef] [PubMed]

45. Golbraikh, A.; Tropsha, A. Predictive QSAR modeling based on diversity sampling of experimental datasets for the training and test set selection. *Mol. Divers.* **2002**, *5*, 231–243. [CrossRef]

46. Daina, A.; Michielin, O.; Zoete, V. iLOGP: A simple, robust, and efficient description of n-octanol/water partition coefficient for drug design using the GB/SA approach. *J. Chem. Inf. Model.* **2014**, *54*, 3284–3301. [CrossRef] [PubMed]

47. Cheng, T.; Zhao, Y.; Li, X.; Lin, F.; Xu, Y.; Zhang, X.; Li, Y.; Wang, R.; Lai, L. Computation of octanol-water partition coefficients by guiding an additive model with knowledge. *J. Chem. Inf. Model.* **2007**, *47*, 2140–2148. [CrossRef]

48. Wildman, S.A.; Crippen, G.M. Prediction of physicochemical parameters by atomic contributions. *J. Chem. Inf. Comput. Sci.* **1999**, *39*, 868–873. [CrossRef]

49. Moriguchi, I.; HIRONO, S.; LIU, Q.; NAKAGOME, I.; MATSUSHITA, Y. Simple method of calculating octanol/water partition coefficient. *Chem. Pharm. Bull.* **1992**, *40*, 127–130. [CrossRef]

50. Sanders, M.P.; Barbosa, A.J.; Zarzycka, B.; Nicolaes, G.A.; Klomp, J.P.; de Vlieg, J.; Del Rio, A. Comparative analysis of pharmacophore screening tools. *J. Chem. Inf. Model.* **2012**, *52*, 1607–1620. [CrossRef]

51. Daina, A.; Zoete, V. A BOILED-Egg to predict gastrointestinal absorption and brain penetration of small molecules. *ChemMedChem* **2016**, *11*, 1117–1121. [CrossRef]

52. *PyMOL*, version 1.8; Schrödinger LLC.: New York, NY, USA, 2015.

53. Sanner, M.; Huey, R.; Dallakyan, S.; Karnati, S.; Lindstrom, W.; Morris, G.; Norledge, B.; Omelchenko, A.; Stoffler, D.; Vareille, G. *AutoDockTools*, version 1.4.5.; The Scripps Research Institute: La Jolla, CA, USA, 2007.

54. *Discovery Studio Visualizer*; Dassault Systemes BIOVIA: San Diego, CA, USA, 2017; p. 936.

55. Trott, O.; Olson, A.J. AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J. Comput. Chem.* **2009**, *31*, 455–461. [CrossRef]

56. Van Der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A.E.; Berendsen, H.J. GROMACS: Fast, flexible, and free. *J. Comput. Chem.* **2005**, *26*, 1701–1718. [CrossRef]

57. Lindorff-Larsen, K.; Piana, S.; Palmo, K.; Maragakis, P.; Klepeis, J.L.; Dror, R.O.; Shaw, D.E. Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins* **2010**, *78*, 1950–1958. [CrossRef]

58. Sousa da Silva, A.W.; Vranken, W.F. ACPYPE—AnteChamber PYthon Parser interfacE. *BMC Res. Notes* **2012**, *5*, 367. [CrossRef]

59. Berendsen, H.J.C.; Postma, J.P.M.; van Gunsteren, W.F.; DiNola, A.; Haak, J.R. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* **1984**, *81*, 3684–3690. [CrossRef]

60. Murugan, N.A.; Kumar, S.; Jeyakanthan, J.; Srivastava, V. Searching for target-specific and multi-targeting organics for Covid-19 in the Drugbank database with a double scoring approach. *Sci. Rep.* **2020**, *10*, 19125. [CrossRef]

61. Kumari, R.; Kumar, R.; Lynn, A. g_mmpbsa–a GROMACS tool for high-throughput MM-PBSA calculations. *J. Chem. Inf. Model.* **2014**, *54*, 1951–1962. [CrossRef]

62. García-Jacas, C.s.R.; Marrero-Ponce, Y.; Cortés-Guzmán, F.; Suárez-Lezcano, J.; Martinez-Rios, F.O.; García-González, L.A.; Pupo-Meriño, M.; Martinez-Mayorga, K. Enhancing acute oral toxicity predictions by using consensus modeling and algebraic form-based 0D-to-2D molecular encodes. *Chem. Res. Toxicol.* **2019**, *32*, 1178–1192. [CrossRef]

63. Roy, K.; Das, R.N.; Ambure, P.; Aher, R.B. Be aware of error measures. Further studies on validation of predictive QSAR models. *Chemom. Intell. Lab. Syst.* **2016**, *152*, 18–33. [CrossRef]

64. Martin, Y.C. A bioavailability score. *J. Med. Chem.* **2005**, *48*, 3164–3170. [CrossRef]

65. Carrano, A.; Snkhchyan, H.; Kooij, G.; van der Pol, S.; van Horssen, J.; Veerhuis, R.; Hoozemans, J.; Rozemuller, A.; de Vries, H.E. ATP-binding cassette transporters P-glycoprotein and breast cancer related protein are reduced in capillary cerebral amyloid angiopathy. *Neurobiol. Aging* **2014**, *35*, 565–575. [CrossRef]

66. Daina, A.; Michielin, O.; Zoete, V. SwissADME: A free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules. *Sci. Rep.* **2017**, *7*, 42717. [CrossRef]

67. Delaney, J.S. ESOL: Estimating aqueous solubility directly from molecular structure. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 1000–1005. [CrossRef]

68. Ali, J.; Camilleri, P.; Brown, M.B.; Hutt, A.J.; Kirton, S.B. Revisiting the general solubility equation: In silico prediction of aqueous solubility incorporating the effect of topographical polar surface area. *J. Chem. Inf. Model.* **2012**, *52*, 420–428. [CrossRef] [PubMed]

69. Yang, L.P.H.; Keating, G.M. Fenofibric Acid. *Am. J. Cardiovasc. Drugs* **2009**, *9*, 401–409. [CrossRef]

70. Moen, M.D. Indacaterol. *Drugs* **2010**, *70*, 2269–2280. [CrossRef] [PubMed]

71. Gupta, S.C.; Patchva, S.; Aggarwal, B.B. Therapeutic Roles of Curcumin: Lessons Learned from Clinical Trials. *AAPS J.* **2013**, *15*, 195–218. [CrossRef]

72. Verta, R.; Grange, C.; Gurrieri, M.; Borga, S.; Nardini, P.; Argenziano, M.; Ghe, C.; Cavalli, R.; Benetti, E.; Miglio, G.; et al. Effect of Bilastine on Diabetic Nephropathy in DBA2/J Mice. *Int. J. Mol. Sci.* **2019**, *20*, 2554. [CrossRef] [PubMed]

73. Pinna, A.; Blasetti, F.; Ricci, G.D.; Boscia, F. Bromfenac eyedrops in the treatment of diabetic macular edema: A pilot study. *Eur. J. Ophthalmol.* **2017**, *27*, 326–330. [CrossRef]