

In silico approaches to explore structure of new GPR 119 agonists for treatment of type 2 diabetes mellitus

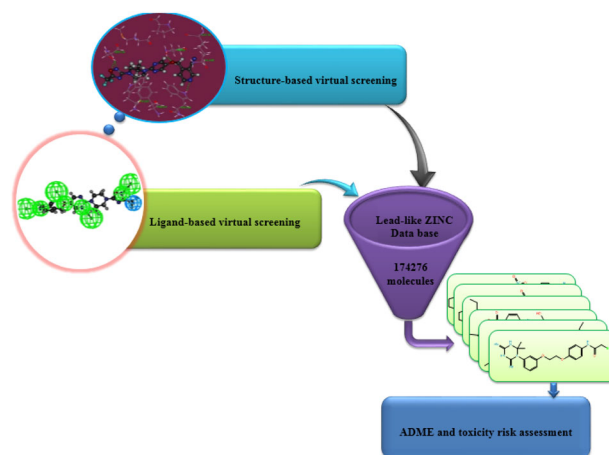
Fereshteh Shiri¹ · Maryam Teymoori¹

Received: 14 October 2016 / Accepted: 13 February 2017 / Published online: 27 February 2017
© Springer Science+Business Media New York 2017

Abstract The G protein coupled receptor 119 is an exciting and promising target for the treatment of type 2 diabetes mellitus. In this study, three dimensional quantitative structure activity relationships i.e. a comparative molecular field analysis and comparative molecular field analysis region focusing have been carried out on a novel series of G protein coupled receptor 119 agonists. A quality pharmacophore model was built by Hip-Hop algorithm. The best pharmacophore model indicated which feature of hydrogen-bond acceptor and hydrophobic existed around active compounds. As the quality of the pharmacophore model is satisfactory, it was utilized to search the ZINC database for a virtual screening task. The obtained compounds subjected to Lipinski filter at first, were docked with Gold algorithm to discover potent hits. The energy difference between the highest occupied molecular orbital and lowest unoccupied molecular orbital (“gap”) implies high reactivity of the most active molecule in the active site of protein. In addition, the molecular electrostatic potential energy at density functional theory level confirms the results from molecular docking. In silico absorption, distribution, metabolism, and excretion and toxicity risk assessment analysis were carried out on the seven hits with the highest Gold score fitness. Six of the new hits from virtual screening had diverse structures and are reported as new

scaffold candidates for G protein coupled receptor 119 agonists.

Graphical Abstract



Keywords G protein coupled receptor 119 · Pharmacophore · Molecular docking · Virtual screening · ADMET

Introduction

Type 2 diabetes mellitus (T2DM) is one of the most common diseases which many of the world's people suffer from, and is increasing rapidly. T2DM constitutes approximately 90% diabetic patients (Sharma and Chandola 2011). It is estimated that 382 million people have diabetes and this number is expected to reach 592 million by 2035 (Guariguata et al. 2014). Therefore, there is a great

Electronic supplementary material The online version of this article (doi:10.1007/s00044-017-1808-y) contains supplementary material, which is available to authorized users.

✉ Fereshteh Shiri
fereshteh.shiri@gmail.com

¹ Department of Chemistry, University of Zabol, P.O. Box 98615-538, Zabol, Iran

requirement for the discovery of novel drugs in this field. Intensive efforts have been made worldwide to develop therapies for type 2 diabetes. The attractive targeting of G-protein-coupled receptor 119 (GPR119) with the small molecules inhibitors is one of the most promising new approaches for the treatment of type 2 diabetes and other elements of the metabolic syndrome. GPR119 is a class A type receptor (Shah 2009) that is expressed in pancreatic β -cells and incretin releasing cells in the gastrointestinal (GI) tract (Overton et al. 2008). One of reasons that cause T2DM is gradual loss of beta-cells function in the pancreatic (Wellenzohn et al. 2012). Agonists of GPR119 as anti-diabetics attract interest for their capability to simulate an incretin response in the gut and glucose-dependent insulin release in the pancreas (McClure et al. 2011).

In recent years, the computer-aided drug design approaches (Cern et al. 2014) have made a large contribution to the design of novel drugs. Comparative molecular field analysis (CoMFA) (Cramer et al. 1988; Shiri et al. 2015) studies through interpretation of contour maps have given insight to design the new inhibitors. The pharmacophore approach (Ambure et al. 2014) (Ambure et al. 2014; Pirhadi and Ghasemi 2012) is effective in drug design. According to a definition by IUPAC, a pharmacophore model is ‘an ensemble of steric and electronic features that is necessary to ensure the optimal supramolecular interactions with a specific biological target and to trigger (or block) its biological response’ (Yang 2010). Virtual screening as a power tool to introduce novel hits have been used (Cano et al. 2014; Shoichet 2004). Molecular docking (Gschwend et al. 1996) is applied for the investigation of the interactions between protein and ligand and to attain bioactive conformers (Ghasemi et al. 2013; Li et al. 2011; Negoro et al. 2012). In the present work, we are to find new inhibitors for type 2 diabetes from the approaches of pharmacophore modeling, molecular docking, and virtual screening.

Moreover the electronic effects play a key role in drug discovery process (Feng et al. 2005). Herein, docking studies and quantum chemistry calculations were performed on the basis of density functional theory (DFT) theory in order to explore some amino acids involved in the binding site by paying special attention to the electronic effects which are related to highest occupied molecular orbital (HOMO) and lowest unoccupied molecular orbital (LUMO) energies and molecular electrostatic potential (MEP) map. Herein, we have introduced the novel hits with different scaffolds using virtual screening in the ZINC 3D database which involve ligand-based pharmacophore and molecular docking. Finally, absorption, distribution, metabolism, and excretion (ADME) prediction and toxicity risk assessment analysis were carried out by virtual screening to obtain the required ADMET properties.

Material and method

Data set

A series of 44 small molecules of GPR119 agonist and their functional potency (EC_{50} value) were taken from the literature (Scott et al. 2012). The dataset was split to training set (33 compound) and test set (10 compound) by Kennard and Stone algorithm to maximize the diversity of the test set and to test the predictive accuracy of the model when extrapolating outside the training set (Kennard and Stone 1969). Due to large residual compound **34** is outlier and was removed from the training set. The EC_{50} values (μM), (The EC_{50} is the concentration of a drug that gives half-maximal response) was converted to pEC_{50} in the molar range in order to give reliable numerical values. The structure of the compounds and their functional potency are shown in Tables 1a–e.

Table 1a Structures and experimental biological activities (pEC_{50}) for compounds of the general structure I

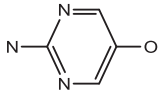
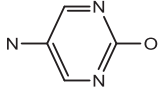
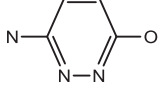
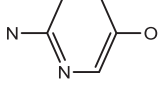
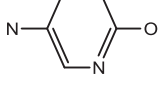
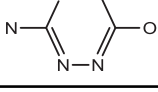
Compound	Core	EC_{50} (μM)	pEC_{50}
1		0.065	7.19
2		1.084	5.96
3		0.485	6.31
4		0.186	6.73
5		0.737	6.13
6		0.161	6.79

Table 1b Structures and experimental biological activities (pEC₅₀) for compounds of the general structure II

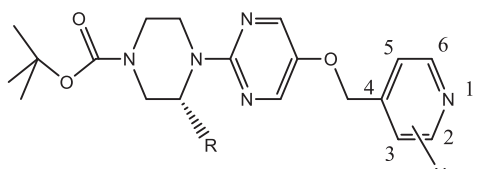
Compound	Core	EC ₅₀ (μM)	pEC ₅₀
7		0.893	6.05
8		0.041	7.39
9		0.019	7.72
10		0.054	7.27
11		0.054	7.27
12		0.466	6.33
13		0.064	7.19
14		0.039	7.41
15		2.217	5.65

Table 1b continued

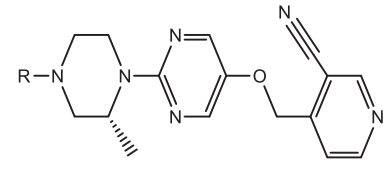
Compound	Core	EC ₅₀ (μM)	pEC ₅₀
16		0.156	6.81

Table 1c Structures and experimental biological activities (pEC₅₀) for compounds of the general structure III

Compound	R	EC ₅₀ (μM)	pEC ₅₀
17		0.029	7.54
18		0.079	7.1
19		0.039	7.41
20		0.006	8.22
21		0.003	8.52
22		0.016	7.8
23		0.045	7.35

Table 1d Structures and experimental biological activities (pEC₅₀) for compounds of the general structure IV


Compound	R	X	EC ₅₀ (μM)	pEC ₅₀
24	H	H	0.621	6.21
25	H	3-Me	0.075	7.12
26	H	3-OMe	1.925	5.72
27	H	3-F	0.063	7.2
28	H	3-CN	0.021	7.68
29	H	3,5-F,F	0.154	6.81
30	Me	3-Me	0.184	6.74
31	Me	3-CN	0.005	8.3

Table 1e Structures and experimental biological activities (pEC₅₀) for compounds of the general structure V


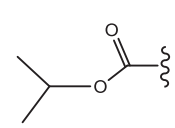
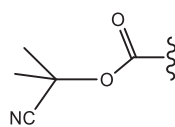
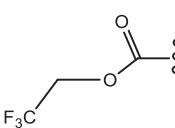
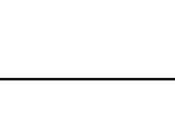
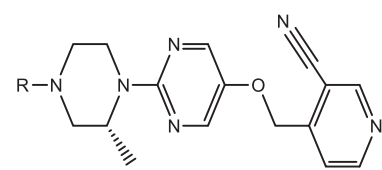
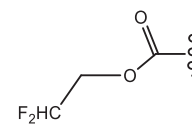
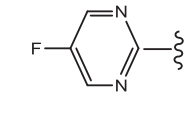
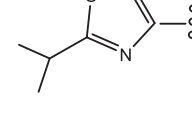
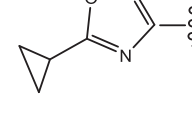
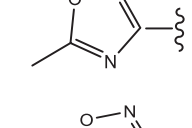
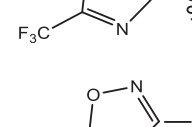
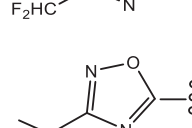
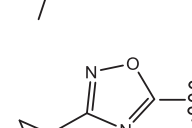
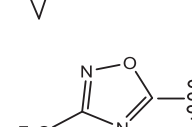
Compound	R	EC ₅₀ (μM)	pEC ₅₀
32		0.112	6.95
33		0.054	7.27
34		0.020	7.7
35		0.100	7

Table 1e continued


Compound	R	EC ₅₀ (μM)	pEC ₅₀
36		0.063	7.2
37		0.008	8.1
38		0.014	7.85
39		0.234	6.63
40		0.002	8.7
41		0.021	7.68
42		0.010	8
43		0.025	7.6
44		0.005	8.3

Molecular modeling and alignment

The 3D structures of compounds **1–44** (Tables 1a–e) were built in SYBYL 7.3 molecular modeling software package from Tripos, Inc., St. Louis, MO. The partial atomic charges of structures were calculated by Gasteiger–Hückel and their energy was minimized using the Tripos force field with a distance-dependent dielectric and the Powell conjugate gradient algorithm convergence criterion of 0.01 kcal/mol (Clark et al. 1989). Alignment is a critical step in CoMFA modeling. Appropriate alignment method caused to acquire a model with high quality statistical parameters. Here, we applied rigid body alignment by the SYBYL Distill module. Compound **40** was considered as a template for the alignment of compounds because it had the highest potency. Alignment of molecules was performed with maximum common substructure by distill method without including bond types in rings (Pirhadi and Ghasemi 2012; Pirhadi et al. 2014). Because CoMFA models are very sensitive to the different space orientations of the molecular aggregate with regard to the grid box, all-orientation search (AOS) strategy was used to achieve the best orientation and minimize the effect of the initial orientation of aligned compounds. AOS module was written in SYBYL programming language (Wang et al. 1998).

CoMFA study

The CoMFA model described the molecular characterization with van der Waals and electrostatic fields, using Lennard–Jones and Coulombic potential, respectively. The CoMFA descriptor fields were calculated at each grid point by a sp^3 hybridized carbon atom with a charge of +1 as the probe atom in a 3D cubic lattice. Column filtering was set to 2.0 kcal/mol. Cut-off for both steric and electrostatic field was set to 30 kcal/mol. Also, in method comparative molecular field analysis region focusing (CoMFA-RF) applies weights to the lattice points which attenuate the contribution of these points to subsequent analysis and refining model, here ‘StDev* Coefficients’ values and different weighing factors were applied in addition to grid spacing for achieving the better models (Pirhadi and Ghasemi 2010; Shiri et al. 2016). Partial least squares (PLS) regression was used to find the relationship between the CoMFA descriptors as independent variables and the biological activities as dependent variables (Wold et al. 2001).

Model validation

Leave-one-out cross-validation (LOO-CV) as internal validation was employed to validate the derived quantitative structure activity relationships (QSAR) model. In the LOO-CV method, one compound of the training set was excluded, and a new QSAR model was built by the remaining compounds and

the activity of the excluded compound was predicted using the new constructing model. This process was repeated until all compounds of the training set were excluded once. The optimum number principle component for the model as the number leading to the highest value of q^2 is defined as

$$q^2 = 1 - \frac{\sum_{i=1}^n (Y_{\text{pred}} - Y_{\text{obs}})^2}{\sum_{i=1}^n (Y_{\text{obs}} - Y_{\text{mean}})^2}$$

where Y_{obs} , Y_{pred} , and Y_{mean} are the observed, predicted, and mean of the biological activity values, respectively, and n is the number of compounds in the training set. A high q^2 value ($q^2 > 0.5$) was used as the evidence of high predictive ability of the model (Gibbs 2000; Hawkins et al. 2003). Golbarikh and Tropsha reported that the high value of q^2 is essential and important but not adequate for a predictive model (Lu et al. 2010) and an external analysis should be employed by a test set of molecules to investigate the prediction of model. They also stated that a QSAR model is predictive if it fulfills the following conditions (Tropsha et al. 2003):

$$q^2 \geq 0.5$$

$$R^2 \geq 0.6$$

$$\frac{(R^2 - R_0^2)}{R^2} < 0.1 \quad \text{or} \quad \frac{R^2 - R_0'^2}{R^2} < 0.1$$

$$0.85 \leq k \leq 1.15 \quad \text{or} \quad 0.85 \leq k' \leq 1.15$$

Where R^2 is the squared correlation coefficient values between the observed and predicted values of the test set compounds (Lu et al. 2010). R_0^2 and $R_0'^2$ are squared correlation coefficient values for predicted vs. observed and observed vs. predicted activities, through origin, respectively, and k' is the slope of regression lines through the origin (Lu et al. 2010). In order to evaluate the predictive power of the CoMFA models, the activity values of the external test set of compounds not used in model generation, were predicted. The predictive R^2 (R^2_{pred}) of the external prediction was obtained as following formula:

$$R^2_{\text{pred}} = 1 - \frac{\sum (Y_{\text{obs}} - Y_{\text{pred}})^2}{\sum (Y_{\text{obs}} - \bar{Y}_{\text{mean}})^2}$$

where Y_{obs} , Y_{pred} , and \bar{Y}_{mean} are the observed, predicted, and mean values of the activity, respectively.

Molecular docking

The crystal structure of GPR119 has not been published yet. The pdb homology modeled from the primary sequences of human GPR119 (Q8TDV5) was taken from the CSSB (<http://cssb.biology.gatech.edu/skolnick/sup/GPCRt3>). Among the top five generated homology models, the first one was selected based on the highest C-score (2.1) and used for docking study.

Molecular docking was done in Discovery studio 2.5 package (AccelrysInc, San Diego, CA, USA) with Gold algorithm. The highest active compound **40** was typed with CHARMM force field and the partial charges were determined by the Momany–Rone option. Then, the resulting compound was minimized by a smart minimizer algorithm which performs 1000 steps of the steepest descent with a RMS gradient tolerance of 3, followed by conjugate gradient minimization. The preparation and minimization of protein were made, the protein structure was typed with CHARMM force field, the partial charge was computed by Momany–Rone, hydrogen atoms were added, and all water molecules were removed and pH of protein was adjusted to almost neutral 7.4, using protein preparation protocol. The active site of the target protein was created as a spherical region with a radius of 16 Å which shows atoms of the ligand and the side-chains of the residues of the receptor within 16 Å from the center of the binding site are free to move. Other parameters were fixed by default protocol settings. The most active compound **40** was docked into active site of the protein. The best pose among 10 given poses with the higher Gold score fitness function was chosen in order to achieve the key interaction between the compound and the amino acid of active site.

Quantum chemical calculations

In this study, the geometry optimization and calculations were performed at the DFT level on a personal computer by energy optimization, using the Gaussian 09 (Frisch et al. 2009) program package. The HOMO and the LUMO energies at B3LYP/6-31G (d,p) level were calculated for bioactive conformer obtained of compound **40** in the docking studies. The MEP was also studied at the same level. The MEP surface was created using Molekel (Varetto 2009).

3D-Pharmacophore generation

A pharmacophore is a structural representation of a spatial arrangement of the essential features (hydrogen bond donor, hydrogen bond acceptor, hydrophobic area, and positively and negatively ionized areas) for a set of molecules to interact with the specific target receptor (24–26). The pharmacophore model was generated to extract the common important pharmacophore features among the first five most active compounds by Hip-Hop in the discovery studio 2.5 packages. The best pharmacophore model was obtained by two features including hydrogen bond acceptor and hydrophobic feature which were color coded with green and blue, respectively.

Virtual screening

Virtual screening including a wide variety of computational approaches let us derivate the most active hits from a

massive virtual library. The pharmacophore validated model was used as a 3D virtual screening query for the detection of the potent candidates from ZINC chemical database. At first we selected a random collection 124,276 molecule from ZINC database and built an in-house chemical database. For each compound in the database, a fast conformer generation approach produced 100 conformers allowing a maximum energy of 15 kcal/mol above the global energy minimum for the conformation searching. In the first screening step, a total 7430 compounds were detected by pharmacophore model. The hit compounds were further screened by pharmacokinetic property filters, Lipinski's rule of five to make them more drug-like because the molecules with molecular weight (MW) more than 500 calculated log P higher than 5, and hydrogen-bond donors, or hydrogen-bond acceptors more than 5 and 10 respectively, were rejected from the database (32, 33). Based on Lipinski's rules, 264 compounds were retrieved. The obtained hit compounds were docked into the protein structure as a structure-based screening using a fast screening method, GOLD, within Discovery Studio 2.5 package. Finally, seven compounds were chosen at this step using a high Gold fitness score. The virtual screening steps are shown in a flowchart in Fig. 1.

ADME and toxicity risk assessment studies

The best binding interactions with a target for the best drug are necessary but not a sufficient condition. An ideal oral medicine absorbed rapidly and completely from the digestive tract, and distributed exactly and specially to its target in the body, does not metabolize the way that immediately clear its activity, and eliminate in a proper way without

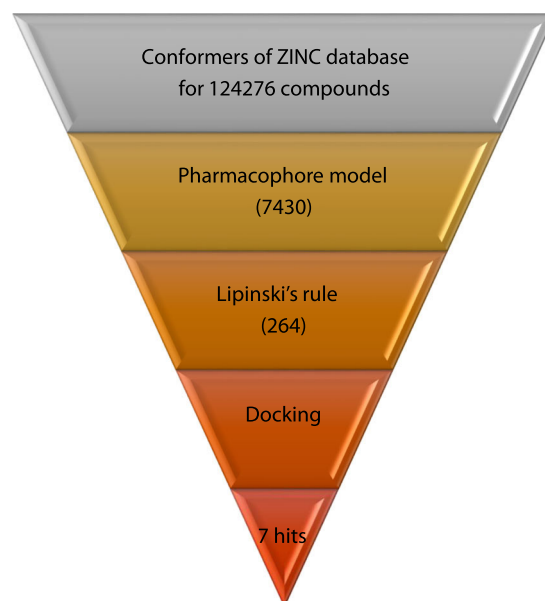


Fig. 1 The flowchart of the combined virtual screening

Table 2 Summary of the statistical parameters for the CoMFA and CoMFA-RF models

Statistical parameters	CoMFA	CoMFA- RF
q^2	0.509	0.605
R^2_{ncv}	0.944	0.945
R^2_{pred}	0.77	0.88
R^2_{bs}	0.969	0.965
SEE	0.205	0.204
SEP	0.214	0.198
F	72.52	72.86
R^2	0.774	0.851
R^2_o	0.757	0.850
$R_o'^2$	0.762	0.835
$R^2 - R_o'^2/R^2$	0.022	0.001
$R^2 - R_o'^2/R_o'^2$	0.016	0.018
k	0.861	0.956
k'	0.987	0.989
n	4	6

q^2 cross-validated correlation coefficient, R^2_{ncv} non-cross-validated correlation coefficient, n number of optimum component, R^2_{pred} predictive correlation coefficient, R^2_o (test) correlation coefficient for regression through the origin for predicted vs. observed activities, $R_o'^2$ (test) correlation coefficient for regression through origin for observed vs. predicted activities

leading harm. The ADME are the main subjects in pharmacokinetics (Hodgson 2001). The ADME properties prediction of compounds before synthesis would be beneficial in the process of drug discovery. The chemists need this data to improve the pharmacokinetic features of compounds. Some chemical descriptors such as MW (less MW for better absorption) were used to measure the pharmacokinetic parameters. Drugs with more MW include many polar functional groups and rotatable bonds so their absorption is poor, and drugs with low MW can result in the extracellular solution of body organs rather quickly (Gurantra 2001). Lipinski rules (Lipinski et al. 2001) were used as a filter to assess the medicinal ability of the compounds in virtual screening step. Some of computed pharmacokinetic parameters were blood-brain barrier (logBB), apparent Caco-2 permeability, apparent MDCK permeability, logKp for skin permeability, the distribution volume, and plasma protein binding (logK_{h_{sa}}) which make effect for the distribution of a drug in the body (Reichel and Begley 1998). LogP (the octanol–water partition coefficient) is a physicochemical parameter of a drug's hydrophobicity property, and also logP and MW are effective in drug excreting from the body. It is obvious that compounds with greater lipophilicity have low absorption and more metabolisms in liver since they can bind macromolecules easier causing their toxicity, also fast renal clearance is related to hydrophilic and small compounds. On the other hand, the metabolism of

Table 3 Gold fitness, experimental and predicted pEC₅₀ values with residual values of the training and test set compounds

Compound	Experimental	CoMFA		CoMFA-RF		Gold fitness
		Pred.	Res.	Pred.	Res.	
1	7.19	7.43	-0.24	7.55	-0.36	49.72
2	5.96	5.89	0.07	5.98	-0.02	50.66
3	6.31	6.52	-0.21	6.4	-0.09	51.66
4^a	6.73	6.49	0.24	6.49	0.24	50.48
5^a	6.13	6.02	0.11	5.93	0.2	50.02
6	6.79	6.74	0.05	6.81	-0.02	49.34
7	6.05	6.25	-0.2	6.07	-0.02	48.98
8^a	7.39	7.39	0	7.3	0.09	51.84
9	7.72	7.39	0.33	7.3	0.42	51.84
10^a	7.27	7.21	0.06	7.16	0.11	52.04
11	7.27	7.21	0.06	7.16	0.11	52.04
12	6.33	6.26	0.07	6.2	0.13	51.83
13	7.19	7.21	-0.02	7.21	-0.02	50.58
14	7.41	7.38	0.03	7.4	0.01	51.53
15	5.65	5.63	0.02	5.68	-0.03	50.73
16	6.81	6.75	0.06	6.86	-0.05	52.03
17^a	7.54	7.39	0.15	7.6	-0.06	54.14
18	7.1	7.14	-0.04	7.18	-0.08	52.38
19	7.41	7.39	0.02	7.6	-0.19	49.03
20^a	8.22	7.72	0.5	7.7	0.52	54.79
21	8.52	8.6	-0.08	8.42	0.1	53.58
22	7.8	7.96	-0.16	8.1	-0.3	49.57
23	7.35	7.34	0.01	7.19	0.16	50.02
24	6.21	6.34	-0.13	6.41	-0.2	46.44
25	7.12	6.93	0.19	6.96	0.16	47.47
26	5.72	5.68	0.04	5.7	0.02	45.14
27	7.2	6.86	0.34	6.76	0.44	47.4
28	7.68	7.92	-0.24	7.77	-0.09	47.71
29	6.81	6.8	0.01	6.79	0.02	49.7
30	6.74	6.87	-0.13	7.08	-0.34	52.48
31	8.3	7.86	0.44	7.96	0.34	47.86
32	6.95	7.24	-0.29	7.01	-0.06	46.09
33^a	7.27	7.83	-0.56	7.59	-0.32	44.59
34	7.7	6.79	0.91	7.07	0.63	51.06
35	7	6.95	0.05	7.04	-0.04	47.92
36	7.2	7.32	-0.12	7.29	-0.09	52.78
37	8.1	7.75	0.35	7.96	0.14	51.48
38^a	7.85	7.57	0.28	7.71	0.14	55.13
39	6.63	6.84	-0.21	6.68	-0.05	51.05
40	8.7	8.57	0.13	8.59	0.11	55.99
41	7.68	7.63	0.06	7.66	0.02	52.26
42^a	8	7.69	0.31	7.82	0.18	55.18
43^a	7.6	7.8	-0.2	7.88	-0.28	54.15
44	8.3	8.56	-0.26	8.45	-0.15	56.03

^a Prediction set

most medicines in the liver is related to hydrophobic and big compounds (Lombardo et al. 2003; Testa et al. 2001). The in-silico prediction of ADME descriptors were calculated using QikProp v.3.0 program (Jorgensen 2006) and were investigated about related standard ranges (<http://www.schrodinger.com/QikProp/>). All hits were subjected to the toxicity risk assessment (mutagenicity, tumorigenicity, irritant, and reproductive effects) using OSIRIS program. Also, their drug-likeness drug score values were determined.

Results and discussion

CoMFA analysis

The statistical parameters from CoMFA and CoMFA-RF models constructed with steric and electrostatic fields are mentioned in Table 2. The best results were obtained for CoMFA-RF model. LOO-CV procedure of the CoMFA-RF model showed a high q^2 value of 0.605 with 6 latent variables. The non-cross-validated PLS analysis of the model leading to $R^2_{\text{pred}} = 0.88$ and standard error of estimation (SEE) of 0.204, $F = 72.86$. To evaluate the robustness and statistical confidence of the models, a total of 100 runs bootstrapping analysis was carried out. A high R^2_{bs} (0.965) indicated the degree of confidence in analysis. These results detected a good conventional statistical correlation. The observed and predicted activities, based on the prediction of the models are listed in Table 3. Figure 2 depicts the relationship between the experimental and predicted activities based on the derived QSAR models. The interpretation of this model's contour maps is easier and more rational than the corresponding CoMFA model.

The steric (*green* and *yellow*) and electrostatic (*blue* and *red*) fields of CoMFA-RF model as 3D contour maps for visualizing the results are exhibited in Fig. 3a, b for compound 40 as template. The changes in fields around the molecules were associated with the changes in volume and position of the contour maps leads to the differences in biological activity. Higher activity values are related to greater bulky group near *green*, minor bulky group near *yellow*, more electrons donating near *blue* and more electron withdrawing near *red* regions in all grid potential analysis.

The *yellow* contour near R2 substituent demonstrated the activity of compound 21 because the presence of *F* substituent on the aryl ring is higher than compound 22 with cyanide group here. Also, compound 18 with more bulky group has lower inhibitory activity than compound 19. A favorable steric green contour covering R4 substituent and showing bulky groups at this position is able to increase the activity. This can be expressed by comparing the order of activities in these compounds: 33 > 32 and 38 > 39

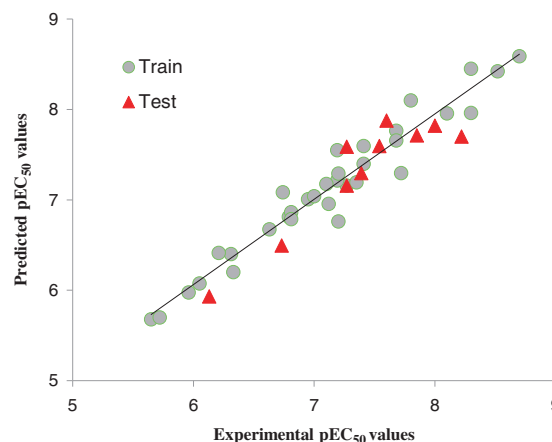


Fig. 2 Predicted against observed activities for training and test sets of GPR 119 agonists based on the CoMFA model

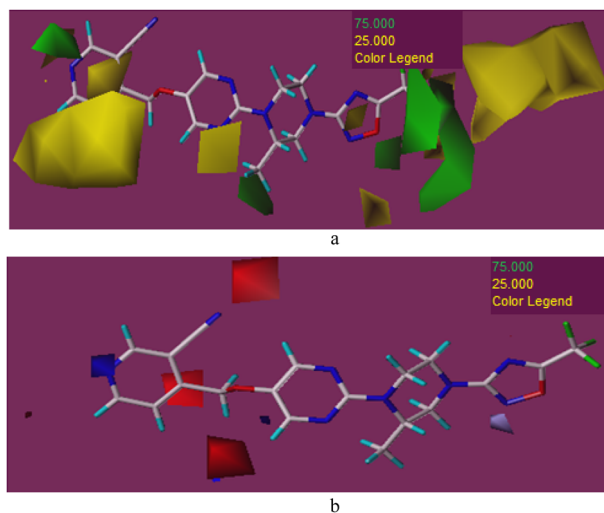


Fig. 3 Contour maps of **a** steric and **b** electrostatic fields of CoMFA based on the most active compound 40

(compound 38 has cyclopropyl group instead of methyl group in compound 39 at R4 position). Two small *red* contours near X substituent at C5 position of the ring shows presence of the electronegative groups are beneficial for the activity. As it is obvious, the compound 29 with *F* electronegative substituent is more active than the compounds 24 and 26 with H atom. The presence of small *blue* contour around the R2 substituent expressed that electropositive groups in this area will increase the activity. For example, the activity of compound 20 ($pEC_{50} = 8.22$) (N groups as resonance in aromatic ring) is better than compound 19 ($pEC_{50} = 7.41$).

Pharmacophore modeling analysis

Using ligand-based pharmacophore modeling, first, common features were obtained by five compounds 20, 21, 31,

40, and **44** as active compounds with the Hip-Hop algorithm. Hydrogen-bond acceptor, hydrogen-bond acceptor lipid, hydrophobic, hydrophobic aromatic, and ring aromatic features were chosen as the initial chemical features in the qualitative pharmacophore generation. The qualitative pharmacophore model indicated which feature of Hydrogen-bond acceptor and hydrophobic existed around active compounds. In brief, the chemical features common to a set of active compounds were estimated by the program and some hypotheses were generated for their activity. Spatial dispositions of pharmacophore features providing the compounds' relative alignment in the active site of the target are hypotheses. The top ranked solutions, HYP1 (rank = 82.95), had two features; hydrogen bond acceptor and hydrophobic feature. The pharmacophore model which was mapped to the most active compound **40** is shown in Fig. 4. For model validation, 110 GPR 119 agonists associated with EC50 obtained from binding database (<http://www.bindingdb.org>) which is a large library. Among these compounds, nearly 90 compounds had the two pharmacophore features of the best pharmacophore model. This fact let us use this pharmacophore model for virtual screening. The initial pharmacophore model (HYP1) returned 90 molecules (82% of the database, Table 1S in Supplementary Information). These results validate the HYP1 pharmacophore model as a screening methodology for potent inhibitors of T2DM.

Molecular docking analysis

Molecular docking computations were carried out to discover the better binding conformations of the inhibitors. Many interactions between the more active compounds and amino acids of the active site were observed. The amino acids of the active site are shown in the literature (Negoro et al. 2012). The key interactions between the ligands and the active site confirm the three dimensional quantitative structure activity relationships (3D-QSAR) model results as well. Binding modes of the most active ligand (compound **40**) with receptor were investigated and the best docked conformation of this compound is shown in Fig. 5. The Gold program validated docking performance using GOLDScore fitness and Binding energy (kcal/mol) which showed a high reliability of the docking method. Molecular docking study between the most active compound and the receptor indicated hydrogen bonding interactions with three active site residues of Thr86 (O H–N, bond length = 3.2 Å), Trp238 (O H–O, bond length = 2.47 Å), and Ser156 (O....H–O, bond length = 2.32 Å and N.....H–O, bond length = 1.98 Å). Also, there are several van der Waals interactions between Ile136, Ala90, Trp265, and Phe241 residues and more active molecule in the dataset. Furthermore, all 44 compounds associated their

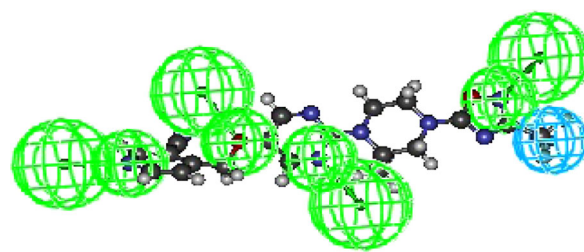


Fig. 4 Mapping of the most active compound **40** on the pharmacophore model. These chemical features are color coded with *green*, hydrogen-bond acceptor and *blue*, hydrophobic feature

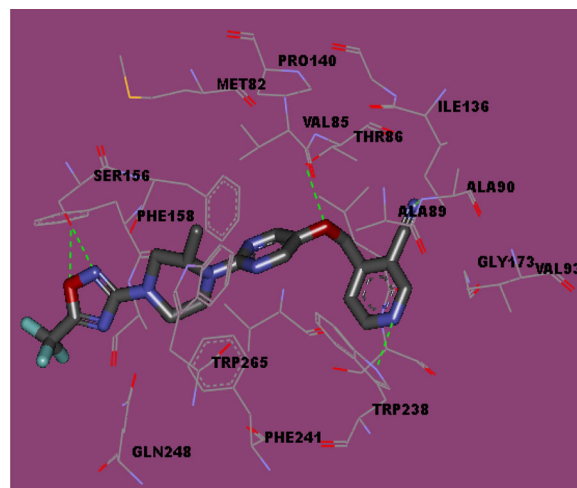


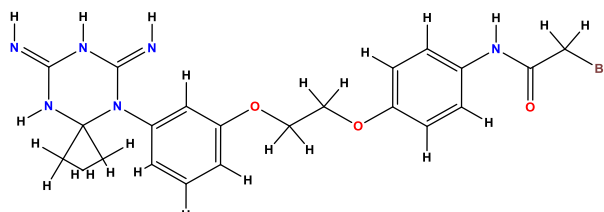
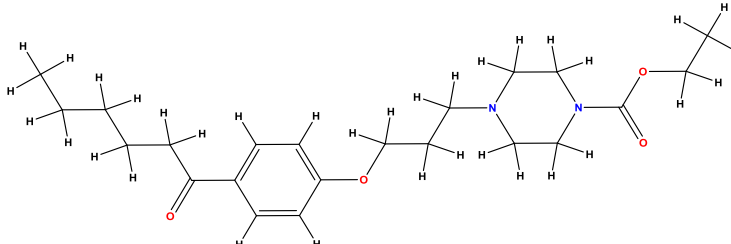
Fig. 5 The best docked conformation of the most active compounds (compound **40**), in the binding site of GPR119. Hydrogen bonding interactions as highlighted by the dashed lines in green color formed between compound **40** and HSA

GOLDScore values are shown in Table 4. Site-directed mutagenesis of target residues was applied to validation of GPR119 homology modeled. We selected three key residues (THR86, SER156, and TRP238) of interest from the docking area that had H-bond with compound **40** for the site-directed mutagenesis. Then, single, double, and triple mutated of 3 key residues to ALA were done. Compound **40** was docked to all mutants GPR119. Table 2S clearly depicted that all mutant GPR119 had small GOLDScore values than the GPR119 without mutagenesis. Also, there was not any H-bond in docking results of all mutant GPR119 and compound **40**.

Frontier molecular orbital

According to the frontier molecular orbitals theory, HOMO and LUMO energy are two significant indicators of chemical reactivity. The character electron donor and electron acceptor of compound were measured by the HOMO and LUMO energies, respectively. The energy difference between the HOMO and LUMO (gap) is an important factor

Table 4 continued

Hits	ZINC ID	Structures	Gold fitness
Vs6	ZINC112531		53.57
Vs7	ZINC22132		53.56

to represent a simple measure of molecule stability. A small gap value implies high reactivity of molecules in reactions while a large gap value implies high stability of molecules and low reactivity in reactions.

$$\text{GAP} = E_{\text{HOMO}} - E_{\text{LUMO}}$$

HOMO represents the ability to donate an electron while LUMO as an electron acceptor represents the ability to obtain an electron. In Fig. 6, HOMO and LUMO orbitals of the conformer obtained from docking with HOMO–LUMO gap are given. As seen in the figure, in the HOMO, the electrons are mainly delocalized on the nitrogen of the piperazine and pyrimidine ring, and partially on the nitrogen and oxygen of 1,2,4-oxadiazole ring; in the LUMO the electrons are completely delocalized on the pyridine and cyanide group. These observations confirm the obtained results from molecular docking.

Molecular electrostatic potential

The MEP, $V(r)$, at a given point $r(x, y, z)$ located in the neighborhood of a molecule can be defined in terms of the interaction energy between the electrical charge generated from the molecule electrons and nuclei, as well as a positive test charge (a proton) placed at r . The $V(r)$ values for the system studied were calculated employing the equation (Politzer and Murray 2002)

$$V(r) = \sum Z_A / |R_A - r| - \rho(r') / |r' - r| d^3 r'$$

where Z_A is the charge of nucleus A located at R_A , $\rho(r')$ is the electronic density function of the molecule, and r' is the dummy integration variable.

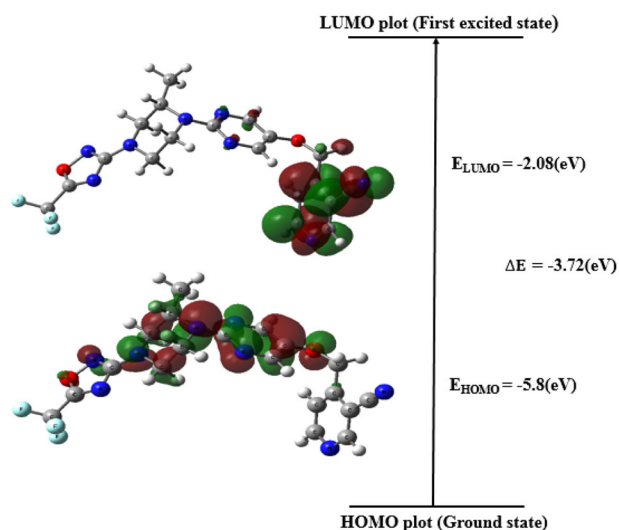


Fig. 6 The HOMO and LUMO orbitals and the energy levels for the most active molecule (40)

The MEP is related to the electronic density and can also be used as a highly beneficial descriptor for the determination of sites for electrophilic attack and nucleophilic reactions as well as hydrogen-bonding interactions (Kaufman 1979; Pomelli et al. 2001). The electrostatic potential $V(r)$ is also well-suited for analyzing processes based on the recognition of one molecule from another one as in drug–receptor and enzyme substrate interactions, since it is through their potentials that the two species first ‘see’ each other (Politzer et al. 1985). Defined as a real physical property, $V(r)$ can be determined experimentally by diffraction or by computational methods (Politzer et al. 1981).

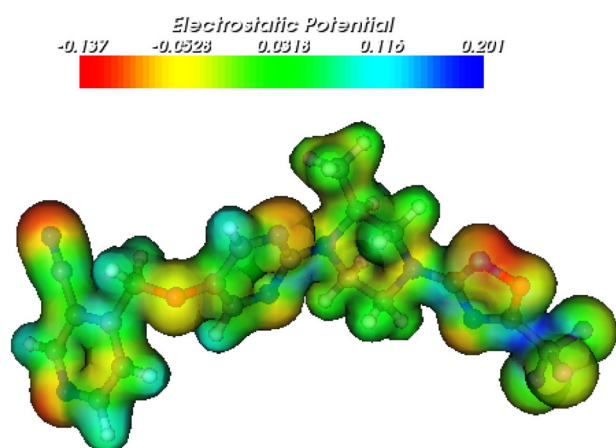


Fig. 7 MEP map (in a.u.) of the most active molecule (**40**)

MEP was calculated at the B3LYP/6-31G (d,p) optimized geometry so that it was possible to anticipate reactive sites for electrophilic and nucleophilic attack for the title molecule. As shown in Fig. 7, the two regions namely negative (*red*) and positive (*blue*) were related to electrophilic and nucleophilic reactivity respectively. As can be seen from the figure, the negative region for electrophilic attack is at pyridine and 1,2,4-oxadiazole ring and oxygen with *red* and *yellow* color. The negative regions are localized on the unprotonated nitrogen atom of the pyridine ring, on the oxygen between the pyridine and pyrimidine rings and on the nitrogen and oxygen of 1,2,4-oxadiazole ring that are H-bond acceptor from backbone NH group of the Trp238, backbone OH group of the Thr86 and backbone OH group of the Ser156, respectively in docking. However, there are some positive regions are localized on hydrogen atoms in pyridine and pyrimidine rings that indicated they are possible sites for nucleophilic attack.

Virtual screening results

The best pharmacophore model was used as a 3D virtual screening query for regaining novel and potent candidates from ZINC chemical database. At first we selected a random collection 124,276 compound from it. The model was filtered using different steps. In first screening step, the compounds of database were obtained by pharmacophore model. A total 7430 compound was retrieved. The compounds of pharmacophore model were screened by pharmacokinetic property filters, Lipinski's rule of five (Lipinski 2000; Lipinski et al. 2012). Based on Lipinski's rule of five 264 compound was retrieved. The next step, the yield hit compounds was docked in to the protein structure as a structure—based on screening using a fast screening method, GOLD, within Discovery Studio 2.5 program package.

The structures of the seven hits associated their GOLDscore values are shown in Table 4. The retrieved compounds show new scaffolds and they are useful for drug design. All of seven retrieved compounds had amine group and aromatic ring in their structures which were mapped to the hydrogen bond acceptor and hydrophobic features of pharmacophore model. In addition, six of seven hits have new scaffold. The first hit with high GOLDscore value had two hydrogen bond with Glu261 and Arg81. One of the retrieved compounds with a GOLDscore of 53.56, namely, **Vs7**, was so similar the dataset compound and it has two hydrogen bond to Thr86 and Arg81. The claim of “discovery of a new scaffold” from virtual screening analysis is supported by having six of seven hits which are not GPR 119 agonists derivatives. We did a search in ChEMBL database (Gaulton et al. 2012) to find the structures retrieved from the virtual screening. Compounds **Vs4** (ChEMBL174261), **Vs6** (ChEMBL419709), and **Vs7** (ChEMBL62803) were in the database. There were several reports of identification **Vs4** (Lipinski 1983), **Vs6** (Andrea and Kalayeh 1991), and **Vs7** (Berlin et al. 2010; Faghieh et al. 2002; Gfesser et al. 2004; Vasudevan et al. 2002) with good potency as inhibitor for different target in literature.

We did a search in ChEMBL database to find the structures retrieved from the virtual screening. Compounds **Vs4** and **Vs7** were in the database. There were two reports inhibitory concentration by compound **Vs7** (i) for Histamine H3 receptor with potency of 8.86 nM (ii) for Histamine H1 receptor with potency of 6.56 nM. Also, we saw some reports inhibitory concentration by compound **Vs4** (i) potency of 2238.7 nM Ubiquitin carboxyl-terminal hydrolase 2 (ii) potency of 4466.8 nM Microtubule-associated protein tau (iii) potency of 5623.4 nM Beta-glucocerebrosidase.

In silico ADME and toxicity risk assessment analysis

In silico ADME prediction was performed on the hits extracted from ZINC (Koes and Camacho 2012) database. For this purpose pharmacokinetic parameters, logBB, skin-permeability coefficient (logKp), apparent Caco-2 and MDCK permeability that the higher the value of MDCK cell, higher the cell permeability, aqueous solubility (log *S*), number of metabolic reactions, log K_{hsa} for serum protein binding, human oral absorption in the GI tract, log P for octanol/water which control hydrophilicity or lipophilicity property of molecules, were calculated. The drug with too much hydrophobic property will be badly soluble in the gut and will solve in fat globules. Hydrophilic compounds unlike hydrophobic ones have low log P , low serum protein binding, potent aqueous solubility and bad cell permeability (Jorgensen 2004). Almost, all hits fell in the acceptable ranges for the properties analyzed for 95% of known drugs, Table 5. Additionally, they all confirm drug-like properties

Table 5 The pharmacokinetic parameters of the retrieved hits compound using Qikprop

Descriptors	108876	114411	32095	35847	102809	112531	22132	Stand. range
Apparent Caco-2 permeability (nm/s)	96	52	352	57	27	230	348	<25 poor, >500 great
	39	27	161	22	30	272	175	<25 poor, >500 great
Apparent MDCK permeability (nm/s)	0.011	0.000	0.110	0.000	0.005	0.002	0.005	mg/cm ² h
	–	–	–4.299	–7.181	–2.881	–5.379	–4.089	–6.5/0.5
Jm (max. transdermal transport rate)	3.943	5.348	89	63	38	88	94	<25% is poor
	58	72	–1.861	–2.610	–2.593	–1.595	–0.698	–3.0/102
logS (aqueous solubility)	–	–	0.209	0.380	–1.065	0.167	0.222	–1.5/105
% human oral absorption in GI ($\pm 20\%$)	2.428	2.868	2.774	2.916	–0.373	3.259	3.762	–2.0/6.5
	–	–	–2.320	–8.164	–5.076	–2.946	–3.781	–8.0 to –1.0, Kp in cm/h
log BB for brain/blood	0.564	0.043						
logK _{hsa} (serum protein binding)	1.382	2.458						
logP for octanol/water	–	–						
Skin-permeability coefficient (logKp)	3.68	4.082						

Table 6 Compliance of the hits with the standard intervals for computational toxicity risk parameters

NO. of hits	Toxicity risks parameters				Drug-likeness parameters (Osiris)		
	MUT	TUM	IRR	REP	CLP	S	DL
1	No risk	No risk	No risk	No risk	0.44	–4.30	–39.38
2	No risk	No risk	No risk	No risk	1.33	–2.26	–1.28
3	No risk	No risk	No risk	No risk	1.36	–5.08	–20.84
4	No risk	No risk	High risk	No risk	4.18	–3.56	–9.54
5	No risk	No risk	No risk	No risk	2.99	–3.13	–4.87
6	No risk	No risk	No risk	No risk	1.87	–4.56	–12.18
7	No risk	No risk	No risk	No risk	–3.63	–1.38	6.15

MUT mutagenicity, *TUM* tumorigenicity, *IRR* irritating effects, *REP* reproductive effects, *CLP* cLogP, *S* solubility, *DL* drug-likeness (calculated through OSIRIS property explorer)

according to Lipinski's rule of 5 (Lipinski 2000; Lipinski et al. 1997), which was second filter applying in the combined virtual screening analysis. All retrieved compounds had amine group in their structures involving effective interactions such as forming hydrogen bonds with receptor. They can pass cell membranes in their non-ionized forms, and they can be solved in water well in their ionized forms (Palm et al. 1997). Toxicity risk assessment parameters and drug relevant properties of hits were within the acceptable interval which valid the hits are as drugs. As it can be seen in Table 6 none of the hits presented a risk of mutagenicity, tumorigenicity, irritating, or reproductive toxicity.

Conclusion

In this study, 3D-QSAR models with high predictive capabilities were derived for a series of GPR119 agonist using

CoMFA method to assist of the design of new potent candidates. Molecular docking was performed to investigate the binding interactions of inhibitors and finding the bioactive conformations. The pharmacophore models have been developed with the aid of Hip-Hop algorithm within Discovery Studio program package. The best pharmacophore model was obtained by two features including hydrogen bond acceptor and hydrophobic feature. The validated pharmacophore model was used as a 3D search query to prescreen the compounds from binding database (<http://www.bindingdb.org>). The Gold program was used as a fast method docking based screening methods. The MEP energy at DFT level confirms the results from molecular docking. These methods allowed us to select seven molecules with high binding affinity for the treatment of type 2 diabetes from the ZINC database. In silico ADME studies indicated the hits to compare the computed values of ADME descriptors with the accepted ranges. Therefore, "Discovery of a new scaffold" and finding the new inhibitors

for the treatment of type 2 diabetes within the hit compounds is possible in near future.

Compliance with Ethical Standards

Conflict of Interest The authors declare that they have no competing interest.

References

- Ambure P, Kar S, Roy K (2014) Pharmacophore mapping-based virtual screening followed by molecular docking studies in search of potential acetylcholinesterase inhibitors as anti-Alzheimer's agents. *Biosystems* 116:10–20
- Andrea T, Kalayeh H (1991) Applications of neural networks in quantitative structure-activity relationships of dihydrofolate reductase inhibitors. *J Med Chem* 34(9):2824–2836
- Berlin M, Boyce CW, de Lera Ruiz M (2010) Histamine H3 receptor as a drug discovery target. *J Med Chem* 54(1):26–53
- Cano G, García-Rodríguez J, Pérez-Sánchez H (2014) Improvement of virtual screening predictions using computational intelligence methods. *Lett Drug Des Discov* 11(1):33–39
- Cern A, Barenholz Y, Tropsha A, Goldblum A (2014) Computer-aided design of liposomal drugs: In silico prediction and experimental validation of drug candidates for liposomal remote loading. *J Control Release* 173:125–131
- Clark M, Cramer RD, Van Opdenbosch N (1989) Validation of the general purpose Tripos 5.2 force field. *J Comput Chem* 10(8):982–1012
- Cramer RD, Patterson DE, Bunce JD (1988) Comparative molecular field analysis (CoMFA). 1. Effect of shape on binding of steroids to carrier proteins. *J Am Chem Soc* 110(18):5959–5967
- Faghhi R, Dwight W, Gentles R, Phelan K, Esbenshade TA, Ireland L, Miller TR, Kang CH, Fox GB, Gopalakrishnan SM, Hancock AA (2002) Structure-activity relationships of non-imidazole H₃ receptor ligands. Part 1. *Bioorg Med Chem Lett* 12(15):2031–2034
- Feng S, Wang Z, He X, Zheng S, Xia Y, Jiang H, Tang X, Bai D (2005) Bis-huperzine B: highly potent and selective acetylcholinesterase inhibitors. *J Med Chem* 48(3):655–657
- Frisch M, Trucks G, Schlegel HB, Scuseria G, Robb M, Cheeseman J, Petersson G (2009) Gaussian 09, revision A. 02, vol 19. Gaussian, Inc., Wallingford, p 227–238
- Gaulton A, Bellis LJ, Bento AP, Chambers J, Davies M, Hersey A, Light Y, McGlinchey S, Michalovich D, Al-Lazikani B, Overington JP (2012) ChEMBL: a large-scale bioactivity database for drug discovery. *Nucleic Acids Res* 40(D1):D1100–D1107
- Gfesser GA, Zhang H, Dinges J, Fox GB, Pan JB, Esbenshade TA, Yao BB, Witte D, Miller TR, Kang CH, Krueger KM (2004) Structure-activity relationships of non-imidazole H₃ receptor ligands. Part 3: 5-Substituted 3-phenyl-1, 2, 4-oxadiazoles as potent antagonists. *Bioorg Med Chem Lett* 14(3):673–676
- Ghasemi JB, Aghaee E, Jabbari A (2013) Docking, CoMFA and CoMSIA studies of a series of N-Benzoylated phenoxazines and phenothiazines derivatives as antiproliferative agents. *Bull Korean Chem Soc* 34(3):899–906
- Gibbs JB (2000) Mechanism-based target identification and drug discovery in cancer research. *Science* 287(5460):1969–1973
- Gschwend DA, Good AC, Kuntz ID (1996) Molecular docking towards drug discovery. *J Mol Recogn* 9(2):175–186
- Guariguata L, Whiting D, Hambleton I, Beagley J, Linnenkamp U, Shaw J (2014) Global estimates of diabetes prevalence for 2013 and projections for 2035. *Diabetes Res Clin Pr* 103(2):137–149
- Guranatra C (2001) Drug metabolism and pharmacokinetics in drug discovery: a primer for bioanalytical chemists, part I. *Curr Sep* 19(1):17–24
- Hawkins DM, Basak SC, Mills D (2003) Assessing model fit by cross-validation. *J Chem Inf Model* 43(2):579–586
- Hodgson J (2001) ADMET-turning chemicals into drugs. *Nat Biotechnol* 19(8):722–726
- Jorgensen W (2004) The many roles of computation in drug discovery. *Science* 303(5665):1813–1818
- Jorgensen W (2006) QikProp, version Please provide city name in the reference Jorgensen (2006). 3.0. Schrodinger, LLC, New York
- Kaufman JJ (1979) Quantum chemical and physicochemical influences on structure-activity relations and drug design. *Int J Quantum Chem* 16(2):221–241
- Kennard RW, Stone LA (1969) Computer aided design of experiments. *Technometrics* 11(1):137–148
- Koes DR, Camacho CJ (2012) ZINCPharmer: pharmacophore search of the ZINC database. *Nucleic Acids Res* 40(W1):W409–W414
- Li XB, Wang SQ, Xu WR, Wang RL, Chou KC (2011) Novel inhibitor design for hemagglutinin against H1N1 influenza virus by core hopping method. *PLoS One* 6(11):1–6
- Lipinski CA (1983) Bioisosteric design of conformationally restricted pyridyltriazole histamine H₂-receptor antagonists. *J Med Chem* 26(1):1–6
- Lipinski CA (2000) Drug-like properties and the causes of poor solubility and poor permeability. *J Pharmacol Toxicol Methods* 44(1):235–249
- Lipinski CA, Lombardo F, Dominy BW, Feeney PJ (1997) Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv Drug Deliver Rev* 23(1–3):3–25
- Lipinski CA, Lombardo F, Dominy BW, Feeney PJ (2001) Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv Drug Deliv Rev* 23(1–3):3–25
- Lipinski CA, Lombardo F, Dominy BW, Feeney PJ (2012) Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv Drug Deliv Rev* 64:4–17
- Lombardo F, Gifford E, Shalaeva MY (2003) *In silico* ADME prediction: data, models, facts and myths. *Mini-Rev Med Chem* 3(8):861–875
- Lu P, Wei X, Zhang R (2010) CoMFA and CoMSIA 3D-QSAR studies on quionolone carboxylic acid derivatives inhibitors of HIV-1 integrase. *Eur J Med Chem* 45(8):3413–3419
- McClure KF, Darout E, Guimarães CR, DeNinno MP, Mascitti V, Munchhof MJ, Robinson RP, Kohrt J, Harris AR, Moore DE, Li B (2011) Activation of the G-protein-coupled receptor 119: a conformation-based hypothesis for understanding agonist response. *J Med Chem* 54(6):1948–1952
- Negoro K, Yonetoku Y, Moritomo A, Hayakawa M, Iikubo K, Yoshida S, Takeuchi M, Ohta M (2012) Synthesis and structure-activity relationship of fused-pyrimidine derivatives as a series of novel GPR119 agonists. *Bioorg Med Chem* 20(21):6442–6451
- Overton H, Fyfe M, Reynet C (2008) GPR119, a novel G protein-coupled receptor target for the treatment of type 2 diabetes and obesity. *Br J Pharmacol* 153(S1):S76–S81
- Palm K, Stenberg P, Luthman K, Artursson P (1997) Polar molecular surface properties predict the intestinal absorption of drugs in humans. *Pharmaceutical Research* 14(5):568–571
- Pirhadi S, Ghasemi JB (2010) 3D-QSAR analysis of human immunodeficiency virus entry-1 inhibitors by CoMFA and CoMSIA. *Eur J Med Chem* 45(11):4897–4903

- Pirhadi S, Ghasemi JB (2012) Pharmacophore identification, molecular docking, virtual screening, and in silico ADME studies of non-nucleoside reverse transcriptase inhibitors. *Mol Inform* 31(11–12):856–866
- Pirhadi S, Shiri F, Ghasemi JB (2014) Pharmacophore elucidation and 3D-QSAR analysis of a new class of highly potent inhibitors of acid ceramidase based on maximum common sub-structure and field fit alignment methods. *J Iran Chem Soc* 11(5):1329–1336
- Politzer P, Kammeyer CW, Bauer J, Hedges WL (1981) Polar properties of carbon monoxide. *J Phys Chem* 85(26):4057–4060
- Politzer P, Laurence P, Jayasuriya K (1985) Structure-activity correlation in mechanism studies and predictive toxicology. *Environ Health Persp* 61:191–202
- Politzer P, Murray JS (2002) The fundamental nature and role of the electrostatic potential in atoms and molecules. *Theor Chim Acta* 108(3):134–142
- Pomelli CS, Tomasi J, Barone V (2001) An improved iterative solution to solve the electrostatic problem in the polarizable continuum model. *Theor Chem Acc* 105(6):446–451
- Reichel A, Begley DJ (1998) Potential of immobilized artificial membranes for predicting drug penetration across the blood–brain barrier. *Pharm Res* 15(8):1270–1274
- Scott JS, Birch AM, Brocklehurst KJ, Broo A, Brown HS, Butlin RJ, Clarke DS, Davidsson O, Ertan A, Goldberg K, Groombridge SD (2012) Use of small-molecule crystal structures to address solubility in a novel series of G protein coupled receptor 119 agonists: optimization of a lead and in vivo evaluation. *J Med Chem* 55(11):5361–5379
- Shah U (2009) GPR119 agonists: a promising new approach for the treatment of type 2 diabetes and related metabolic disorders. *Curr Opin Drug Discov Devel* 12(4):519–532
- Sharma H, Chandola H (2011) Ayurvedic concept of obesity, metabolic syndrome, and diabetes mellitus. *J Altern Complement Med* 17(6):549–552
- Shiri F, Bakhshayesh S, Ghasemi JB (2015) Computer-aided molecular design of (E)-N-Aryl-2-ethene-sulfonamide analogues as microtubule targeted agents in prostate cancer. *Arab J Chem*. doi:10.1016/j.arabjc.2014.11.063
- Shiri F, Rakhshani-morad S, Samzadeh-kermani A, Karimi P (2016) Computer-aided molecular design of some indolinone derivatives of PLK4 inhibitors as novel anti-proliferative agents. *Med Chem Res* 25(11):2643–2665
- Shoichet BK (2004) Virtual screening of chemical libraries. *Nature* 432(7019):862–865
- Testa B, van der Waterbeemd H, Folkers G, Guy R, van de Waterbeemd H (2001) Pharmacokinetic optimization in drug research. Wiley-VCH: Wiley-VCH, Zurich
- Tropsha A, Gramatica P, Gombar VK (2003) The importance of being earnest: validation is the absolute essential for successful application and interpretation of QSPR models. *QSAR Comb Sci* 22(1):69–77
- Varetto U (2009) Molekel 5.4. 0.8. Swiss National Supercomputing Centre, Manno, Switzerland
- Vasudevan A, Conner SE, Gentles RG, Faghieh R, Liu H, Dwight W, Ireland L, Kang CH, Esbenshade TA, Bennani YL, Hancock AA (2002) Synthesis and evaluation of potent pyrrolidine H₃ antagonists. *Bioorg Med Chem Lett* 12(21):3055–3058
- Wang R, Gao Y, Liu L, Lai L (1998) All-orientation search and all-placement search in comparative molecular field analysis. *J Mol Model* 4(8):276–283
- Wellenzohn B, Lessel U, Beller A, Isambert T, Hoenke C, Nosse B (2012) Identification of new potent GPR119 agonists by combining virtual screening and combinatorial chemistry. *J Med Chem* 55(24):11031–11041
- Wold S, Sjöström M, Eriksson L (2001) PLS-regression: a basic tool of chemometrics. *Chemometr Intell Lab* 58(2):109–130
- Yang SY (2010) Pharmacophore modeling and applications in drug discovery: challenges and recent advances. *Drug Discov Today* 15(11):444–450