J Mol Model (2012) 18:2795–2804
DOI 10.1007/s00894-011-1303-1

ORIGINAL PAPER

A specific pharmacophore model of sodium-dependent glucose co-transporter 2 (SGLT2) inhibitors

Chunlei Tang · Xiaoyun Zhu · Dandan Huang · Xin Zan · Baowei Yang · Ying Li · Xiaoyong Du · Hai Qian · Wenlong Huang

Received: 30 September 2011 / Accepted: 4 November 2011 / Published online: 27 November 2011

Abstract Sodium-dependent glucose co-transporter 2 (SGLT2) plays a pivotal role in maintaining glucose equilibrium in the human body, emerging as one of the most promising targets for the treatment of diabetes mellitus type 2. Pharmacophore models of SGLT2 inhibitors have been generated with a training set of 25 SGLT2 inhibitors using Discovery Studio V2.1. The best hypothesis ($Hypo1_{SGLT2}$) contains one hydrogen bond donor, five excluded volumes, one ring aromatic and three hydrophobic features, and has a correlation coefficient of 0.955, cost difference of 68.76, RMSD of 0.85. This model was validated by test set, Fischer randomization test and decoy set methods. The specificity of $Hypo1_{SGLT2}$ was evaluated. The pharmacophore features of $Hypo1_{SGLT2}$ were different from the best pharmacophore model ($Hypo1_{SGLT1}$) of SGLT1 inhibitors we developed. Moreover, $Hypo1_{SGLT2}$ could effectively distinguish selective inhibitors of SGLT2 from those of SGLT1. These results indicate that a highly predictive and specific pharmacophore model of SGLT2 inhibitors has been successfully obtained. Then $Hypo1_{SGLT2}$ was used as a 3D query to screen databases including NCI and Maybridge for identifying new inhibitors of SGLT2. The hit compounds were subsequently subjected to filtering by Lipinski's rule of five. And several compounds selected from the top ranked hits have been suggested for further experimental assay studies.

Keywords Diabetes mellitus type 2 · HypoGen · Pharmacophore · SGLT1 · SGLT2 · Specific

Introduction

According to the epidemiologic research worldwide in recent decades, diabetes is becoming a common and frequently occurring disease, trending to be a major public health problem. Globally, 380 million people might be suffering from diabetes by 2025 [1]. Diabetes mellitus type 2 (T2DM) accounts for almost 90% of diabetes cases, with the property of insulin resistance and beta-cell dysfunction that induces hyperglycemia [2]. Medical complications associated with T2DM include cardiovascular disease, stroke, nephropathy, retinopathy, renal failure, and amputations of the extremities [3]. Several therapeutic agents are available for monotherapy or combination therapy with different mechanisms to treat diabetics [4]. However, in light of the report by United Kingdom Prevention of Diabetes Study, only 25–50% of T2DM patients are effectively treated by current therapies [5]. The obvious need for new approaches to treat patients with uncontrolled T2DM has prompted continuous exploration of alternative targets involved in maintenance of glucose homeostasis.

In recent years, much attention has been given to sodium-dependent glucose co-transporters (SGLTs), medi-

Chunlei Tang and Xiaoyun Zhu contributed equally to this work

Electronic supplementary material The online version of this article (doi:10.1007/s00894-011-1303-1) contains supplementary material, which is available to authorized users.

C. Tang · X. Zhu · D. Huang · X. Zan · B. Yang · Y. Li · X. Du · H. Qian (✉) · W. Huang (✉)
Centre of Drug Discovery, State Key Laboratory of Natural Medicines, China Pharmaceutical University,
24 Tongjiaxiang,
Nanjing, Jiangsu 210009, China
e-mail: qianhai24@163.com
e-mail: ydhuangwenlong@126.com

ators of reabsorption of glucose in the human body. Sodium-dependent glucose co-transporter 2 (SGLT2) is a high-capacity, low-affinity transporter expressed selectively in the S1 domain of the proximal tubule in the kidney and is responsible for 90% of renal glucose reuptake. Sodium-dependent glucose co-transporter 1 (SGLT1), on the other hand, is a low-capacity, high-affinity transporter distributed in the kidney, gut and other tissues, responsible for the remaining 10% of glucose reuptake [6]. Inhibition of SGLTs induces glucose excretion in urine and thereby reduces plasma glucose concentrations [7, 8]. This mechanism is different from the currently available diabetic therapies that target insulin resistance and insulin deficiency, providing a glucose-dependent and insulin-independent pathway to control hyperglycemia. SGLTs have emerged as a very promising approach to the pathophysiologic treatment of T2DM [4, 7, 9]. Although the inhibition of ubiquitous expressed SGLT1 plays a contributing role, current evidence suggests that inhibition of SGLT1 should produce gastrointestinal disturbances [10, 11]. For instance, the O-arylglucoside natural product phlorizin, which is a non-selective SGLT inhibitor, has long been known to cause glucosuria in animals and humans [12]. Therefore, developing selective SGLT2 inhibitors with minimum adverse effects is urgently needed.

In the present study, we have generated pharmacophore models using Discovery Studio V2.1 (DS) for SGLT2 and SGLT1 inhibitors, respectively. With the aim to obtain the specific pharmacophore model of SGLT2 inhibitors that would provide a hypothetical picture of chemical features responsible for activity, we compared the features between the pharmacophores of SGLT2 and SGLT1 inhibitors and used them as 3D queries for mapping compounds with both SGLT2 and SGLT1 inhibitory activity. Finally, these results indicate that a highly predictive and specific pharmacophore model of SGLT2 inhibitors was obtained. The specific pharmacophore model can be utilized as a predictive tool for estimating biological activity of SGLT2 selective inhibitors through virtual screen or molecular designing on the basis of structure-activity analysis.

Materials and methods

Selection of molecules

A set of 110 different SGLT2 inhibitors and 44 different SGLT1 inhibitors has been collected from different references [1, 13–29]. Of those compounds, 25 SGLT2 inhibitors and 23 SGLT1 inhibitors, which achieve satisfactory diversity in both structural and activity ranges, were selected as the training set for pharmacophore models, respectively. The biological activity values of the selected 25 SGLT2 inhibitors span a range of five orders of magnitude (IC_{50} values ranging from 0.3 to 50000 nM), while the selected 23 SGLT1 inhibitors span six (IC_{50} values ranging from 0.17 to 134000 nM). Chemical structures and experimental IC_{50} values of training set molecules of SGLT2 inhibitors and SGLT1 inhibitors are given in Fig. 1 and Fig. S1 (see Supplementary material), respectively. The remaining 85 SGLT2 inhibitors and 21 SGLT1 inhibitors as the test set are shown in Fig. S2 and Fig. S3 (see Supplementary material), respectively.

Diverse conformation generation

Before starting the pharmacophore generation process, conformation analysis of the molecules was performed using the poling algorithm [30]. The poling algorithm eliminates much of the redundancy in conformation generation and improves the coverage of conformational space. The number of conformers generated for each compound was limited to a maximum of 255 with an energy range of 20 kcal mol^{-1}. The conformers of the training set were generated using the *BEST* conformation model generation method of *diverse conformation generation* protocol implemented in DS, which provided complete and improved coverage of conformational space by performing a rigorous energy minimization and optimizing the conformations in both torsional and cartesian space using the poling algorithm [31, 32].

Generation of the 3D pharmacophore

SGLT2 and SGLT1 inhibitors pharmacophore models were developed using the HypoGen module implemented in DS with the conformers generated for the molecules in the training set, respectively [31]. During pharmacophore hypotheses generation, four kinds of chemical features, including hydrogen bond acceptor (HBA), hydrogen bond donor (HBD), hydrophobic (H) and ring aromatic (RA), were selected. The value for "maximum excluded volumes (EV)" and the uncertainty value for the compound activity were set to 5 and 3, respectively, while other parameters were set to their default values.

Pharmacophore model validation

Validation of a quantitative model is performed in order to determine whether the developed model is able to identify active structures and forecast their activities precisely. The quality of pharmacophore can be validated using test set, Fischer's randomization and decoy set methods.

Test set

This method is used to elucidate whether the generated pharmacophore model is proficient to predict the activities of the compounds other than the training set. The conformation generation of the test set was carried out in a similar way

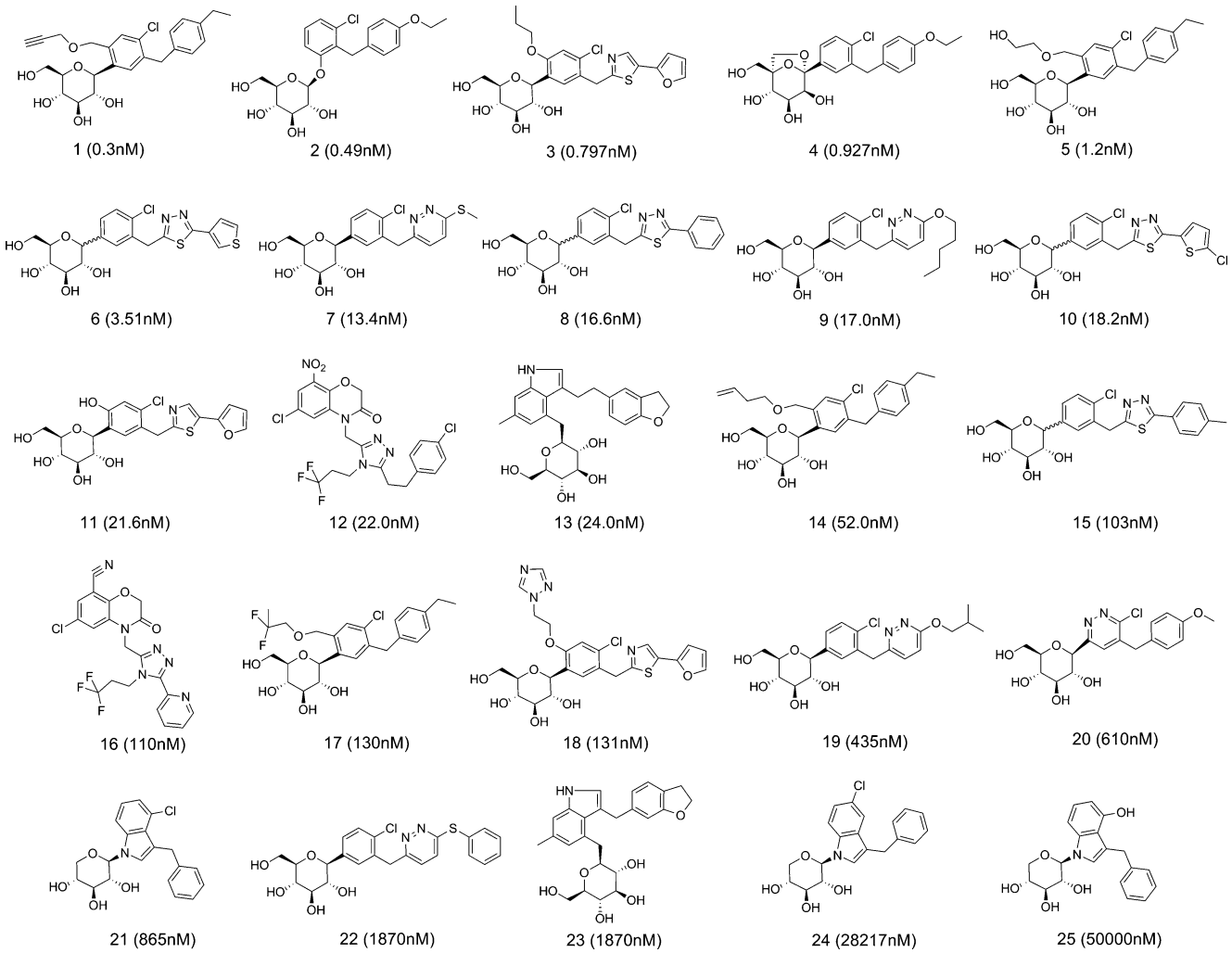

Fig. 1 Chemical structures of 25 SGLT2 inhibitors in training set together with their experimental inhibitory activities (IC_{50} values, in parentheses)

like training set compounds. The conformers were subsequently mapped using the *ligand pharmacophore mapping* protocol with the *best* search option available in DS.

Fischer randomization test

The purpose of this validation is to verify whether there is a good correlation between the chemical structures and biological activities of compounds. The validation was done by generating random spreadsheets for training set molecules, which randomly reassigned activity values to each compound and subsequently generated the hypotheses using the same features and parameters originated for the original pharmacophore hypothesis. To achieve the confidence level of 95%, 19 random spreadsheets were generated. The significance of the hypotheses was calculated using the following formula: $[1-(1+X)/Y]\times 100$, Here, $X=0$ and $Y=(19+1)$, $S=[1-((1+0)/(19+1))]\times 100\% = 95\%$ [33].

Decoy set

The decoy set is used to evaluate the discriminative ability of the best hypothesis by computing goodness of fit score (GF) and enrichment factor (EF). Decoy set contained active compounds of SGLT2 inhibitors and inactive compounds [34]. The screening was performed using the *ligand pharmacophore mapping* protocol implemented in DS, parameters such as total number of compounds in the hit list(Ht), number of active percent of yields, percent ratio of actives in the list, EF, false negatives, false positives and GF were calculated.

Evaluation of the specificity of hypothesis to SGLT2 inhibitors

As it has been mentioned, SGLT2 and SGLT1 are two types of human SGLT proteins. Their amino acid sequence identity reaches 59% [35]. Thus, most inhib-

Table 1 Results of top ten pharmacophore hypotheses of SGLT2 inhibitors generated using training set

	Hypo no.	Total cost	Cost difference[a]	RMSD[b]	Correlation(r)	Features[c]	Max. fit
SGLT2	Hypo1	116.73	68.76	0.85	0.955	HBD, 3H, RA, 5EV	13.64
	Hypo2	120.20	65.28	0.94	0.945	HBA, 3H, RA, 2EV	13.62
	Hypo3	126.37	59.12	1.20	0.908	HBA, 3H, RA	13.14
	Hypo4	129.03	56.46	1.29	0.894	HBA, 3H, RA	13.11
	Hypo5	136.93	48.56	1.57	0.835	HBD, 3H, RA, EV	10.49
	Hypo6	137.34	48.15	1.52	0.847	HBA, 3H, RA	13.20
	Hypo7	139.76	45.72	1.62	0.824	HBD, 3H, RA	11.97
	Hypo8	140.05	45.43	1.61	0.825	HBA, 3H, RA	12.50
	Hypo9	140.15	45.34	1.64	0.817	HBA, 3H, RA	10.83
	Hypo10	140.26	45.23	1.49	0.860	HBA, HBD, H, RA	12.28

[a] "+"indicates that the estimated IC_{50} is higher than the experimental IC_{50}; "-" indicates that the estimated IC_{50} is lower than the experimental IC_{50}

[b] Fit value indicates how well the features in the pharmacophore map the chemical features in the molecule

[c] Activity scale: highly active (IC_{50} < 10 nM, +++), moderately active (10nM ≤ IC_{50} < 1000 nM, ++), and inactive (IC_{50} ≥ 1000 nM, +)

itors targeting against SGLT2 are also likely to target against SGLT1. Furthermore, a specific pharmacophore model of SGLT2 inhibitors can discover some inhibitors selectively targeting against SGLT2. In order to examine the specificity of the pharmacophore model of SGLT2 inhibitors, firstly, compare the chemical features of the best pharmacophore models of SGLT2 and SGLT1 inhibitors; secondly, collect a set of compounds with both SGLT2 and SGLT1 inhibitory activity, subsequently mapped with the best pharmacophore models of SGLT2 and SGLT1 inhibitors, respectively, using the *ligand pharmacophore mapping* protocol available in DS. The specificity was evaluated by the correlation between selectivity and fit values, as well as mapped features.

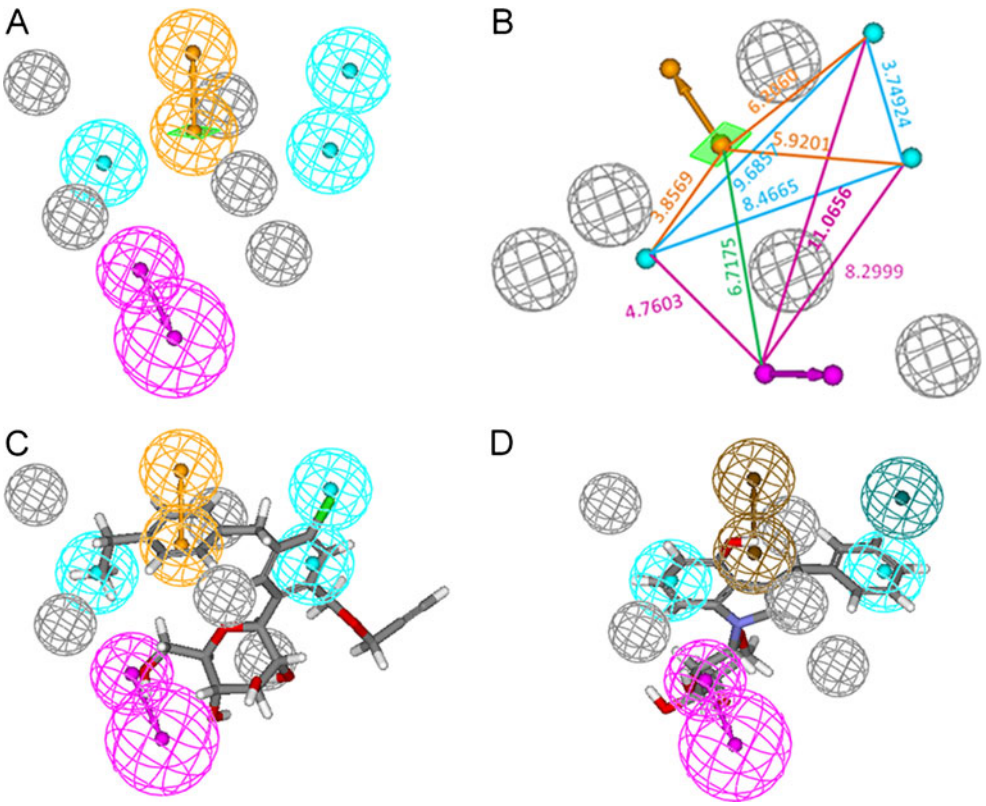

Fig. 2 HypoGen pharmacophore hypothesis for SGLT2 inhibitors. (**a**) The best pharmacophore model $Hypo1_{SGLT2}$. (**b**) 3D spatial relationship and distance constraints of the $Hypo1_{SGLT2}$. (**c**) $Hypo1_{SGLT2}$ aligned to the most active compound in training set. (**d**) $Hypo1_{SGLT2}$ aligned to the most inactive compound in training set. The features are color coded with magenta, hydrogen bond donor; cyan, hydrophobic; orange, ring aromatic; gray, excluded volume

Table 2 Experimental and predicted IC_{50} values of the training set compounds based on the pharmacophore model $Hypo1_{SGLT2}$

Compound no.	Exp. IC_{50} (nM)	Predicted IC50 (nM)	Error[a]	Fit value[b]	Experimental scale[c]	Predicted scale[c]
1	0.3	0.41	+1.4	11.222	+++	+++
2	0.49	0.18	−2.7	11.583	+++	+++
3	0.797	1.07	+1.3	10.811	+++	+++
4	0.927	2.04	+2.2	10.531	+++	+++
5	1.2	5.59	+4.7	10.093	+++	+++
6	3.51	8.79	+2.5	9.896	+++	+++
7	13.4	7.80	−1.7	9.948	++	+++
8	16.6	47.12	+2.8	9.167	++	++
9	17	30.73	+1.8	9.352	++	++
10	18.2	20.12	+1.1	9.536	++	++
11	21.6	5.35	−4.0	10.112	++	+++
12	22	204.06	+9.3	8.53	++	++
13	24	48.09	+2	9.158	++	++
14	52	72.43	+1.4	8.98	++	++
15	103	20.13	−5.1	9.536	++	++
16	110	113.11	+1.02	8.787	++	++
17	130	44.72	−2.9	9.19	++	++
18	131	91.73	−1.4	8.877	++	++
19	435	178.95	−2.4	8.587	++	++
20	610	830.91	+1.4	7.92	++	++
21	865	790.17	−1.1	7.942	++	++
22	1870	1521.01	−1.2	7.658	+	+
23	1870	1841.58	−1.01	7.575	+	+
24	28217	10132.80	−2.8	6.834	+	+
25	50000	15019.30	−3.3	6.66	+	+

[a] "+"indicates that the estimated IC_{50} is higher than the experimental IC_{50}; "-" indicates that the estimated IC_{50} is lower than the experimental IC_{50}

[b] Fit value indicates how well the features in the pharmacophore map the chemical features in the molecule

[c] Activity scale: highly active (IC_{50} < 10 nM, +++), moderately active (10nM ≤ IC_{50} < 1000 nM, ++), and inactive (IC_{50} ≥ 1000 nM, +)

Database screening

The specific HypoGen pharmacophore hypothesis was used as a 3D structural query for retrieving potent molecules from chemical databases including MayBridge and NCI. For each molecule in the database, 255 conformers were generated with the fast conformer generation method, allowing a maximum energy of 20 kcal mol^{-1} above that of the most stable

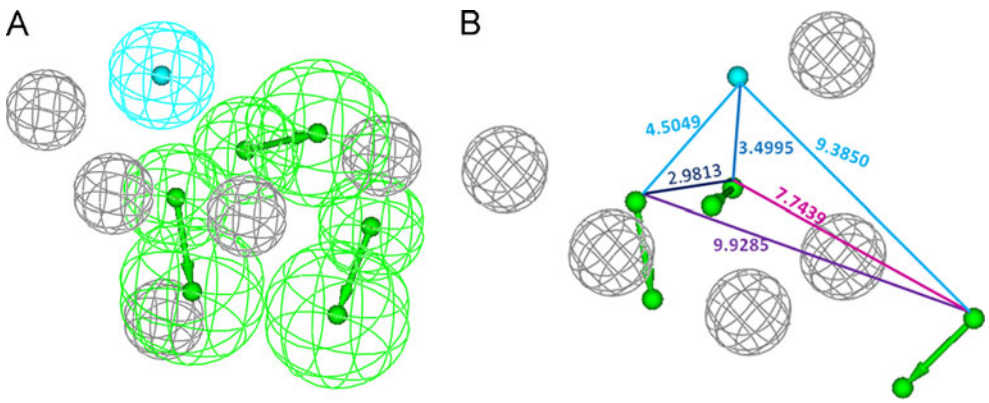

Fig. 3 HypoGen pharmacophore hypothesis for SGLT1 inhibitors. (**a**) The best pharmacophore model $Hypo1_{SGLT1}$. (**b**) 3D spatial relationship and distance constraints of the $Hypo1_{SGLT1}$. The features are color coded with green, hydrogen bond acceptor; cyan, hydrophobic; gray, excluded volume

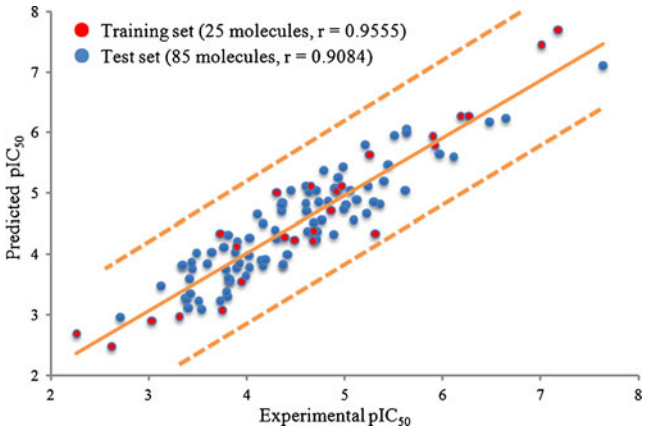

Fig. 4 Plot of the correlation (r) between the experimental activity and the predicted activity by $Hypo1_{SGLT2}$ for 85 test molecules (in cyan) and 25 training set molecules (in red)

conformation. The database screening was carried out using *ligand pharmacophore mapping* protocol implemented in DS with best/flexible search option. The retrieved compounds were filtered by restricting the estimated activity value less than 100 nM and the obtained compounds were further screened in accordance with the Lipinski's rule of five to make them more drug-like.

Results and discussion

Pharmacophore model

Pharmacophore model for SGLT2 inhibitors

Pharmacophores were computed by Hypogen algorithm implemented in DS and top ten hypotheses of SGLT2 inhibitors (Table 1) were exported. The best hypothesis ($Hypo1_{SGLT2}$, Fig. 2a) consisted of one HBD, three H, one RA and five EV features, and has a correlation coefficient of 0.955, cost difference of 68.76, RMSD of 0.85. The fixed and null cost values are 106.187 and 185.485, respectively. The 3D space and distance constraints of these pharmacophore features are shown in Fig. 2b. The most active and inactive compounds in the training set were aligned in $Hypo1_{SGLT2}$ as shown in Fig. 2c, d, respectively. Furthermore, in order to verify the prediction accuracy of $Hypo1_{SGLT2}$, we classified all the training set compounds into three sets based on their activity values: highly active (IC_{50} < 10 nM, +++), moderately active (10nM ≤ IC_{50} < 1000 nM, ++), and inactive (IC_{50} ≥ 1000 nM, +). Table 2 shows the experimental and estimated inhibitory activities of the 25 training set molecules. Obviously, most compounds are correctly predicted except compound 8 (Error: -1.7) and compound 11(Error: -4.0).

Pharmacophore model for SGLT1 inhibitors

Ten hypotheses of SGLT1 inhibitors in Table S1 (see Supplementary material) were exported in the same way as SGLT2 inhibitors. The best hypothesis ($Hypo1_{SGLT1}$, Fig. 3a) consisted of three HBA, one H, and five EV features, and has a correlation coefficient of 0.920, cost difference of 49.25, RMSD of 1.00. The fixed and null cost values are 96.666 and 170.323, respectively. The 3D space and distance constraints of these pharmacophore features are shown in Fig. 3b. Table S2 presents experimental and predicted IC_{50} values of the training set compounds based on the pharmacophore model $Hypo1_{SGLT1}$.

Validation of the pharmacophore model

Prediction with test set molecules

An independent test set_{SGLT2} composed of 85 SGLT2 inhibitors, was used to evaluate the predictive ability of the $Hypo1_{SGLT2}$. Table S3 shows the predicted values of

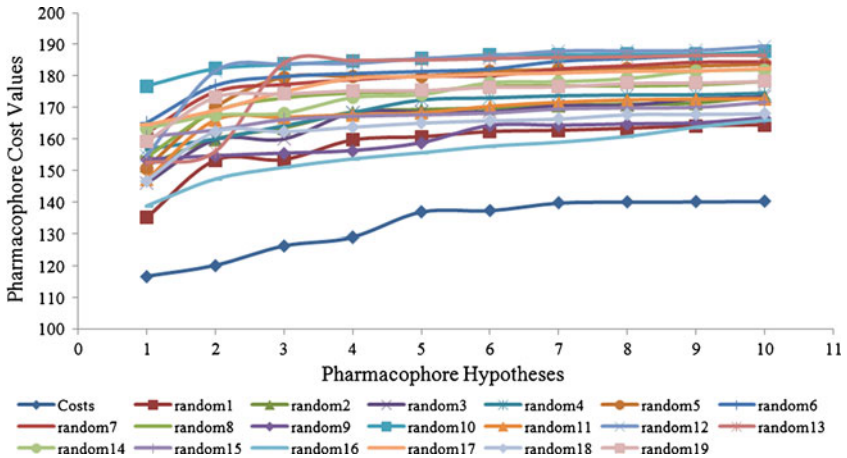

Fig. 5 The difference in costs between HypoGen runs and the scrambled runs. The 95% confidence level was selected

Table 3 Statistical parameter from screening test set molecules

No.	Parameter	Values
1	Total number of molecules in database (D)	459
2	Total number of actives in database (A)	59
3	Total number of hit molecules from the database (Ht)	63
4	Total number of active molecules in hit list (Ha)	56
5	%Yield of actives [(Ha/Ht) × 100]	88.89%
6	%Ratio of actives [(Ha/A) × 100]	94.92%
7	False negatives [A-Ha]	3
8	False positives [Ht-Ha]	7
9	Goodness of fit score[a] (GF)	0.888
10	Enrichment factor[b] (EF)	6.92

[a] [(Ha/4HtA)(3A + Ht) × (1-((Ht-Ha)/(D-A)))]

[b] (Ha × D)/(Ht × A)

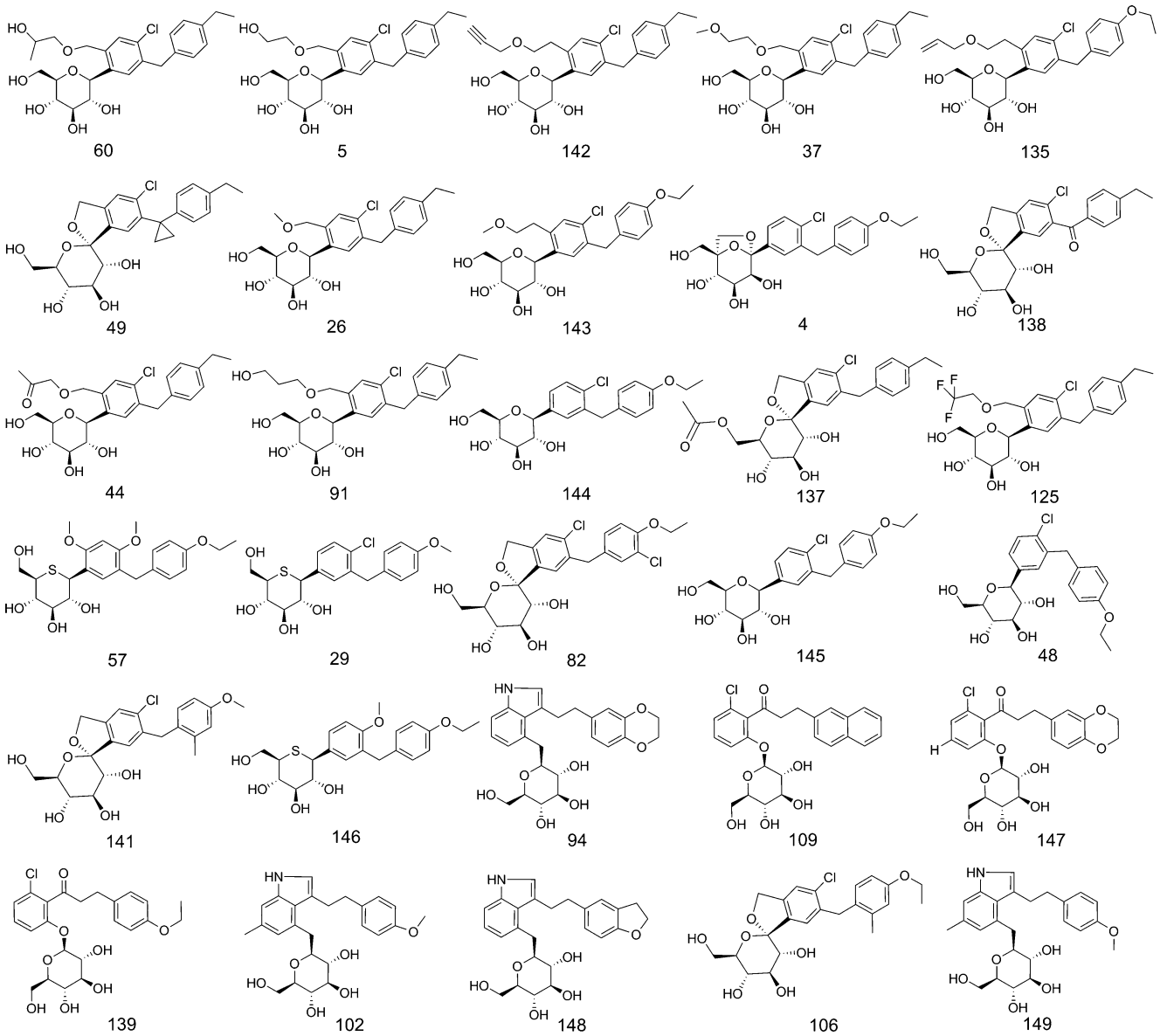

Fig. 6 Chemical structures of 30 compounds with both SGLT2 and SGLT1 inhibitory activity

inhibitory activity of the test set_{SGLT2} compounds together with their corresponding experimental values. The result demonstrates that compounds of all activity scales were predicted properly. Furthermore, a regression analysis (Fig. 4) of the experimental and predicted inhibitory activity values for the test compounds gives a fairly good correlation coefficient (r) of 0.9084. Another independent test set_{SGLT1} composed of 21 SGLT1 inhibitors was used to evaluate the predictive ability of the $Hypo1_{SGLT1}$ (Table S4). The results demonstrate that we have successfully developed two reliable pharmacophore models with high predictivity of $Hypo1_{SGLT2}$ and $Hypo1_{SGLT1}$, respectively.

Fischer randomization test

Fischer randomization test was used to further evaluate the statistical relevance of $Hypo1_{SGLT2}$ and $Hypo1_{SGLT1}$. The test results indicate that $Hypo1_{SGLT2}$ and $Hypo1_{SGLT1}$ are meaningful and successful. Figure 5 clearly exhibits that the $Hypo1_{SGLT2}$ was not generated randomly.

Decoy set

The decoy set of $Hypo1_{SGLT2}$ contains 59 active compounds and 400 inactive compounds. A set of 59 active

Table 4 The experimental IC_{50} values (nM) for both the SGLT2 and SGLT1, selectivity (denoted as the ratio between the IC_{50} values against SGLT2 and SGLT1) and fit values, as well as the mapped features

Compound no.	IC_{50} (nM)		Selectivity	Fit value		Mapped feature[a]				
						$Hypo1_{SGLT2}$			$Hypo1_{SGLT1}$	
	SGLT1	SGLT2	SGLT2/SGLT1	SGLT1	SGLT2	HBD	H	RA	HBA	H
60	60000	16	1/3750	8.626	9.942	+	++	+	-	+
5	4200	1.2	1/3500	8.11	10.097	+	++	+	+	-
142	20000	6.7	1/3250	8.738	8.813	+	++	+	-	+
37	10000	3.6	1/2778	7.329	10.016	++	+	+	-	+
135	31000	12	1/2583	7.139	8.622	+	++	+	-	+
49	>2200	0.9	<1/2222	7.185	11.139	++	++	++	-	+
26	>2200	0.9	<1/2222	9.14	10.2	+	++	+	-	+
143	16000	7	1/2286	6.763	10.246	+	++	+	-	-
4	2050	0.927	1/2211	7.169	10.519	++	++	+	-	+
138	43000	22	1/1955	6.809	10.161	+	+	++	-	-
44	10000	6	1/1667	8.294	9.95	+	+	+	-	+
91	103000	110	1/936	6.567	9.135	+	+	+	-	-
144	1200	1.4	1/857	9.829	9.473	+	+	+	++	-
137	35000	43	1/814	7.21	8.557	+	+	+	-	+
125	134000	170	1/788	7.762	9.898	+	++	+	-	+
57	4270	10.8	1/395	6.429	9.962	+	+	+	-	-
29	260	1.68	1/155	7.613	10.115	+	+	++	-	+
82	10500	68	1/154	6.652	9.24	+	+	+	-	-
145	890	6.7	1/132	9.839	9.473	+	+	+	++	+
48	885	6.7	1/131	9.839	9.473	+	+	+	+	+
141	50000	1000	1/50	6.988	8.507	-	+	+	-	+
146	565	13.4	1/42	7.272	8.599	-	+	+	+	+
94	3610	121	1/30	8.097	8.91	+	+	+	+	-
109	45000	1700	1/26	6.855	7.197	+	-	+	-	+
147	45000	1700	1/26	7.987	8.335	+	-	+	-	+
139	45000	1700	1/26	6.927	9.154	+	+	+	-	+
102	3050	394	1/8	7.477	7.732	-	+	+	-	+
148	145	24	1/6	8.226	7.902	+	-	+	+	+
106	5000	1000	1/5	8.129	8.21	-	+	+	+	-
149	611	163	1/4	7.477	7.732	-	+	+	+	+

[a] "++" means that the compound has mapped the feature very well, "+" means less well, "-"means did not mapped

compounds has been collected from different references [1, 13–29]. The total number of compounds in the hit list (Ht), number of active percent of yields, percent ratio of actives in the hit list, false negatives, false positives, GF and EF are listed in Table 3. The false negatives, false positives, GF and EF are 3, 7, 0.888 and 6.92, respectively, indicating $Hypo1_{SGLT2}$ with high efficiency of screening.

Evaluation of the specificity of $Hypo1_{SGLT2}$ to SGLT2 inhibitors

Comparison of chemical features between $Hypo1_{SGLT2}$ and $Hypo1_{SGLT1}$

As mentioned above, $Hypo1_{SGLT2}$ consisted of one HBD, three H, one RA and five EV features (Fig. 2a, b); on the other hand, $Hypo1_{SGLT1}$ consisted of three HBA, one H, and five EV features (Fig. 3a, b). These results indicate that the pharmacophore models derived from ligands of SGLT2 and SGLT1 are different from each other, which is understandable as it has been demonstrated that the pharmacophore models are intensely dependent on the training set compounds.

The specificity of $Hypo1_{SGLT2}$

We collected a set of 30 compounds (Fig. 6) with both SGLT2 and SGLT1 inhibitory activity. Then for each compound in the compound set, the best fit analysis using $Hypo1_{SGLT2}$ and $Hypo1_{SGLT1}$ was preformed. Table 4 presents the experimental IC_{50} values (nM) for both SGLT2 and SGLT1, selectivity and fit values, as well as mapped features. Clearly, the higher selectivity of SGLT2 inhibitory activity appears the higher fit values with $Hypo1_{SGLT2}$ and lower fit values with $Hypo1_{SGLT1}$ or vice versa. This phenomenon can also be reflected through the mapped features of selective SGLT2 inhibitors. Highly selective SGLT2 inhibitors are mapped with $Hypo1_{SGLT2}$ much better than that of $Hypo1_{SGLT1}$. Based on this analysis, we can conclude that $Hypo1_{SGLT2}$ is the specific pharmacophore model of SGLT2 inhibitors.

Database screening

The validated specific $Hypo1_{SGLT2}$ was used as a 3D structural query for retrieving compounds from MayBridge (59 652 compounds) and NCI (238 819 compounds). As a result, 87 and 326 compounds were retrieved from MayBridge and NCI respectively, with estimated activity value less than 100 nM. These obtained compounds were further refined according to Lipinski's rule of five to make them more drug-like. Finally, a total of 56 compounds passed this filtration. Additionally some of the molecules have been shifted for further experimental assay study.

Conclusions

In this study, chemical features based pharmacophore modeling of SGLT2 inhibitors have been developed using *3D QSAR pharmacophore generation* protocol available in DS 2.1. The best quantitative pharmacophore model, $Hypo1_{SGLT2}$, was characterized by the best correlation coefficient (0.955), the lowest total cost value (116.73), the highest cost difference (68.76), and the lowest RMSD (0.85), and consisted of one HBD, three H, one RA and five EV features. $Hypo1_{SGLT2}$ was further validated by test set, Fischer randomization test and decoy set methods. The test set containing 85 compounds was used in investigating the predictive ability of $Hypo1_{SGLT2}$ and resulted with a correlation coefficient of 0.908, indicating a good predictive capacity. Moreover, other validation methods also have provided reliable results on the strength of $Hypo1_{SGLT2}$. Correspondingly, in order to evaluate the specificity of $Hypo1_{SGLT2}$ to SGLT2 inhibitor, a comparison of chemical features between $Hypo1_{SGLT2}$ and $Hypo1_{SGLT1}$ was made. Furthermore, $Hypo1_{SGLT2}$ and $Hypo1_{SGLT1}$ were also mapped with compounds with both SGLT2 and SGLT1 inhibitory activity. The results clearly demonstrate that $Hypo1_{SGLT2}$ can distinguish selective inhibitors of SGLT2 from those of SGLT1, and the HBD and RA features are likely to be essential to the specificity of $Hypo1_{SGLT2}$. We can conclude that the $Hypo1_{SGLT2}$ truly reflects the features of SGLT2 inhibitors; furthermore, $Hypo1_{SGLT2}$ is also a specific pharmacophore model of SGLT2 inhibitors. Therefore, the specific pharmacophore model of SGLT2 inhibitors should be helpful in identifying novel lead compounds with improved inhibitory activity through 3D database searching, providing valuable tools in the design of new SGLT2 inhibitors.

Acknowledgments The work was supported by the Important National Science & Technology Specific Projects of China (No. 2009ZX09103-033), the National Natural Science Foundation of China (No. 81172932) and the Fundamental Research Funds for the Central Universities of China (No.2 J10023).

References

1. Lee SH, Kim MJ, Kim J, Park HJ, Lee J (2011) Thiazolylmethyl ortho-substituted phenyl glucoside library as novel C-aryl glucoside SGLT2 inhibitors. Eur J Med Chem 46:2662–2675
2. Porte D Jr (2001) Clinical importance of insulin secretion and its interaction with insulin resistance in the treatment of type 2 diabetes mellitus and its complications. Diabetes Metab Res Rev 17:181–188
3. Zhang W, Welihinda A, Mechanic J, Ding H, Zhu L, Lu Y, Deng Z, Sheng Z, Lv B, Chen Y, Roberge JY, Seed B, Wang YX (2011) EGT1442, a potent and selective SGLT2 inhibitor, attenuates blood glucose and HbA(1c) levels in db/db mice and prolongs the survival of stroke-prone rats. Pharmacol Res 63:284–293
4. Washburn WN (2009) Development of the renal glucose reabsorption inhibitors: a new mechanism for the pharmaco-

therapy of diabetes mellitus type 2. J Med Chem 52:1785–1794
5. Group UPDS (1998) Effect of intensive blood-glucose control with metformin on complications in overweight patients with type 2 diabetes (UKPDS 34). UK prospective diabetes study (UKPDS) group. Lancet 352:854–865
6. Kanai Y, Lee WS, You G, Brown D, Hediger MA (1994) The human kidney low affinity Na+/glucose cotransporter SGLT2. Delineation of the major renal reabsorptive mechanism for D-glucose. J Clin Invest 93:397–404
7. Chao EC, Henry RR (2010) SGLT2 inhibition–a novel strategy for diabetes treatment. Nat Rev Drug Discov 9:551–559
8. Neumiller JJ, White JR, Campbell RK (2010) Sodium-glucose co-transport inhibitors: progress and therapeutic potential in type 2 diabetes mellitus. Drugs 70:377–385
9. Nair S, Wilding JPH (2010) Sodium glucose cotransporter 2 inhibitors as a new treatment for diabetes mellitus. J Clin Endocrinol Metab 95:34–42
10. Kasahara M, Maeda M, Hayashi S, Mori Y, Abe T (2001) A missense mutation in the Na+/glucose cotransporter gene SGLT1 in a patient with congenital glucose-galactose malabsorption: normal trafficking but inactivation of the mutant protein. Biochim Biophys Acta 1536:141–147
11. Pascual JM, Wang D, Lecumberri B, Yang H, Mao X, Yang R, De Vivo DC (2004) GLUT1 deficiency and other glucose transporter diseases. Eur J Endocrinol 150:627–633
12. Robinson RP, Mascitti V, Boustany-Kari CM, Carr CL, Foley PM, Kimoto E, Leininger MT, Lowe A, Klenotic MK, MacDonald JI (2010) C-Aryl glycoside inhibitors of SGLT2: exploration of sugar modifications including C-5 spirocyclization. Bioorg Med Chem Lett 20:1569–1572
13. Du X, Lizarzaburu M, Turcotte S, Lee T, Greenberg J, Shan B, Fan P, Ling Y, Medina JC, Houze J (2011) Optimization of triazoles as novel and potent nonphlorizin SGLT2 inhibitors. Bioorg Med Chem Lett 21:3774–3779
14. Dudash J, Zhang X, Zeck RE, Johnson SG, Cox GG, Conway BR, Rybczynski PJ, Demarest KT (2004) Glycosylated dihydrochalcones as potent and selective sodium glucose co-transporter 2 (SGLT2) inhibitors. Bioorg Med Chem Lett 14:5121–5125
15. Ellsworth BA, Meng W, Patel M, Girotra RN, Wu G, Sher PM, Hagan DL, Obermeier MT, Humphreys WG, Robertson JG, Wang A, Han S, Waldron TL, Morgan NN, Whaley JM, Washburn WN (2008) Aglycone exploration of C-arylglucoside inhibitors of renal sodium-dependent glucose transporter SGLT2. Bioorg Med Chem Lett 18:4770–4773
16. Goodwin NC, Mabon R, Harrison BA, Shadoan MK, Almstead ZY, Xie Y, Healy J, Buhring LM, DaCosta CM, Bardenhagen J, Mseeh F, Liu Q, Nouraldeen A, Wilson AG, Kimball SD, Powell DR, Rawlins DB (2009) Novel L-xylose derivatives as selective sodium-dependent glucose cotransporter 2 (SGLT2) inhibitors for the treatment of type 2 diabetes. J Med Chem 52:6201–6204
17. Kakinuma H, Oi T, Hashimoto-Tsuchiya Y, Arai M, Kawakita Y, Fukasawa Y, Iida I, Hagima N, Takeuchi H, Chino Y, Asami J, Okumura-Kitajima L, Io F, Yamamoto D, Miyata N, Takahashi T, Uchida S, Yamamoto K (2010) (1S)-1,5-anhydro-1-[5-(4-ethoxybenzyl)-2-methoxy-4-methylphenyl]-1-thio-D- glucitol (TS-071) is a potent, selective sodium-dependent glucose cotransporter 2 (SGLT2) inhibitor for type 2 diabetes treatment. J Med Chem 53:3247–3261
18. Kang SY, Song KS, Lee J, Lee SH (2010) Synthesis of pyridazine and thiazole analogs as SGLT2 inhibitors. Bioorg Med Chem 18:6069–6079
19. Kim MJ, Lee J, Kang SY, Lee SH, Son EJ, Jung ME, Song KS, Lee M, Han HK, Kim J (2010) Novel C-aryl glucoside SGLT2 inhibitors as potential antidiabetic agents: Pyridazinylmethylphenyl glucoside congeners. Bioorg Med Chem Lett 20:3420–3425
20. Lee J, Lee SH, Seo HJ, Son EJ, Jung ME, Lee M, Han HK, Kim J, Kang J (2010) Novel C-aryl glucoside SGLT2 inhibitors as potential antidiabetic agents: 1,3,4-Thiadiazolylmethylphenyl glucoside congeners. Bioorg Med Chem 18:2178–2194
21. Lv B, Xu B, Feng Y, Peng K, Xu G, Du J, Zhang L, Zhang W, Zhang T, Zhu L, Ding H, Sheng Z, Welihinda A, Seed B, Chen Y (2009) Exploration of O-spiroketal C-arylglucosides as novel and selective renal sodium-dependent glucose co-transporter 2 (SGLT2) inhibitors. Bioorg Med Chem Lett 19:6877–6881
22. Mascitti V, Maurer TS, Robinson RP, Bian J, Boustany-Kari CM, Brandt T, Collman BM, Kalgutkar AS, Klenotic MK, Leininger MT, Lowe A, Maguire RJ, Masterson VM, Miao Z, Mukaiyama E, Patel JD, Pettersen JC, Preville C, Samas B, She L, Sobol Z, Steppan CM, Stevens BD, Thuma BA, Tugnait M, Zeng D, Zhu T (2011) Discovery of a clinical candidate from the structurally unique dioxa-bicyclo[3.2.1]octane class of sodium-dependent glucose cotransporter 2 inhibitors. J Med Chem 54:2952–2960
23. Park EJ, Kong Y, Lee JS, Lee SH, Lee J (2010) Exploration of SAR regarding glucose moiety in novel C-aryl glucoside inhibitors of SGLT2. Bioorg Med Chem Lett 21:742–746
24. Robinson RP, Mascitti V, Boustany-Kari CM, Carr CL, Foley PM, Kimoto E, Leininger MT, Lowe A, Klenotic MK, Macdonald JI, Maguire RJ, Masterson VM, Maurer TS, Miao Z, Patel JD, Preville C, Reese MR, She L, Steppan CM, Thuma BA, Zhu T (2010) C-Aryl glycoside inhibitors of SGLT2: exploration of sugar modifications including C-5 spirocyclization. Bioorg Med Chem Lett 20:1569–1572
25. Washburn WN (2009) Evolution of sodium glucose co-transporter 2 inhibitors as anti-diabetic agents. Expert Opin Ther Pat 19:1485–1499
26. Xu B, Feng Y, Lv B, Xu G, Zhang L, Du J, Peng K, Xu M, Dong J, Zhang W, Zhang T, Zhu L, Ding H, Sheng Z, Welihinda A, Seed B, Chen Y (2010) ortho-Substituted C-aryl glucosides as highly potent and selective renal sodium-dependent glucose co-transporter 2 (SGLT2) inhibitors. Bioorg Med Chem 18:4422–4432
27. Yao CH, Song JS, Chen CT, Yeh TK, Hung MS, Chang CC, Liu YW, Yuan MC, Hsieh CJ, Huang CY, Wang MH, Chiu CH, Hsieh TC, Wu SH, Hsiao WC, Chu KF, Tsai CH, Chao YS, Lee JC (2010) Discovery of novel N-beta-d-Xylosylindole derivatives as sodium-dependent glucose cotransporter 2 (SGLT2) inhibitors for the management of hyperglycemia in diabetes. J Med Chem 54:166–178
28. Zhang X, Urbanski M, Patel M, Cox GG, Zeck RE, Bian H, Conway BR, Beavers MP, Rybczynski PJ, Demarest KT (2006) Indole-glucosides as novel sodium glucose co-transporter 2 (SGLT2) inhibitors. Part 2. Bioorg Med Chem Lett 16:1696–1701
29. Zhang X, Urbanski M, Patel M, Zeck RE, Cox GG, Bian H, Conway BR, Pat Beavers M, Rybczynski PJ, Demarest KT (2005) Heteroaryl-O-glucosides as novel sodium glucose co-transporter 2 inhibitors. Part 1. Bioorg Med Chem Lett 15:5202–5206
30. John S, Thangapandian S, Sakkiah S, Lee KW (2011) Potent BACE-1 inhibitor design using pharmacophore modeling, in silico screening and molecular docking studies. BMC Bioinforma 12 (Suppl 1):S28
31. Inc. A (2010). Discovery studio 21 San Diego
32. Zhu X, Huang D, Lan X, Tang C, Zhu Y, Han J, Huang W, Qian H (2011) The first pharmacophore model for potent G protein-coupled receptor 119 agonist. Eur J Med Chem 46:2901–2907
33. Chen XM, Lu T, Lu S, Li HF, Yuan HL, Ran T, Liu HC, Chen YD (2010) Structure-based and shape-complemented pharmacophore modeling for the discovery of novel checkpoint kinase 1 inhibitors. J Mol Model 16:1195–1204
34. Khalaf RA, Abdula AM, Mubarak MS, Taha MO (2011) Discovery of new β-D-glucosidase inhibitors via pharmacophore modeling and QSAR analysis followed by in silico screening. J Mol Model 17:443–464
35. Wright EM (2001) Renal Na(+)-glucose cotransporters. Am J Physiol Renal Physiol 280:F10–F18