

# Discovery of Novel *N*-Glycoside and Non-Glycoside hSGLT2 Inhibitors for the Treatment of Type 2 Diabetes Mellitus

Chun-Yi Chang<sup>1\*</sup>, Yih Ho<sup>2\*</sup>, Shwu-Jiuan Lin<sup>2</sup>, Hsuan-Liang Liu<sup>4,3</sup>

<sup>1</sup>Institute of Biochemical and Biomedical Engineering, National Taipei University of Technology, Taiwan

<sup>2</sup>School of Pharmacy, Taipei Medical University, Taiwan

<sup>3</sup>Department of Chemical Engineering and Biotechnology, National Taipei University of Technology, Taiwan

Email: f10894@ntut.edu.tw

**How to cite this paper:** Chang, C.-Y., Ho, Y., Lin, S.-J. and Liu, H.-L. (2019) Discovery of Novel *N*-Glycoside and Non-Glycoside hSGLT2 Inhibitors for the Treatment of Type 2 Diabetes Mellitus. *Journal of Diabetes Mellitus*, 9, 77-104.

<https://doi.org/10.4236/jdm.2019.93009>

**Received:** July 25, 2019

**Accepted:** August 19, 2019

**Published:** August 22, 2019

Copyright © 2019 by author(s) and Scientific Research Publishing Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## Abstract

Human sodium-glucose cotransporter 2 (hSGLT2) is a membrane protein responsible for glucose reabsorption from the glomerular filtrate in the proximal tubule. Inhibition of hSGLT2 has been regarded as a brand new therapeutic approach for the treatment of type 2 diabetes mellitus (T2DM) due to its non-insulin related characteristics with less side effects. Current commercially available hSGLT2 inhibitors are all *C*-glycoside inhibitors. Previous studies have reported that *N*-glycoside inhibitors have better potential to serve as new drugs due to their good metabolic stability. In addition, non-glycoside inhibitors have been shown to exhibit the capability to overcome the existing problems of current glycoside inhibitors, including low tissue permeability, poor stability and short serum half-time. Here, we aimed to discover novel *N*-glycoside and non-glycoside hSGLT2 inhibitors by a combination of several computational approaches. A ligand-based pharmacophore model was generated, well validated and subsequently utilized as a 3D query to identify novel hSGLT2 inhibitors from National Cancer Institute (NCI) and Traditional Chinese Medicine (TCM) databases. Finally, one *N*-glycoside (NSC679207) and one non-glycoside (TCM\_Piperenol\_A) hSGLT2 inhibitors were successfully identified, which were proven to exhibit excellent binding affinities, pharmacokinetic properties and less toxicity than the commercially available hSGLT2 inhibitor, canagliflozin, via molecular docking, ADMET prediction, molecular dynamics (MD) simulations and binding free energy calculations. All together, our results strongly suggest that these two compounds have great potential to serve as novel hSGLT2 inhibitors for the treatment of T2DM and their efficacies may be further examined by a series of *in vitro* and/or *in vivo* bioassays.

\*Equal contribution.

---

## Keywords

Human Sodium-Glucose Cotransports 2 (hSGLT2), Type 2 Diabetes Mellitus (T2DM), Ligand-Based Pharmacophore Model, Molecular Docking, Molecular Dynamics (MD) Simulations

---

## 1. Introduction

Diabetes mellitus (DM) has become a severe health problem worldwide and its prevalence has rapidly increased over the past decades [1]. DM is a chronic metabolic disease characterized by the disorder of glucose homeostasis, which further causes long-term hyperglycemia and the development of high-risk vascular complications, including heart failure, nephropathy, retinopathy and neuropathy [2]. There are two common types of DM: type 1 diabetes mellitus (T1DM) and type 2 diabetes mellitus (T2DM). T1DM results from the absolute lack of insulin secretion due to its complete  $\beta$ -cell dysfunction, while T2DM is caused by insulin deficiency and insulin resistance [3]. Nowadays, T2DM accounts for almost 90% - 95% of DM patients and is extending towards a younger population [4]. So far, several commercially available therapeutic agents, such as biguanides, sulfonylureas, thiazolidinediones, DPP-IV inhibitors and GLP-1 receptor agonist, have been proven to exhibit hypoglycemic effects for treating T2DM by increasing insulin secretion or insulin sensitivity [5]. However, not all T2DM patients can be effectively treated by current therapies and these therapeutic agents are also limited by several adverse side effects, including weight gain, hypoglycemia, gastrointestinal flatulence, diarrhea, edema and lactic acidosis [6] [7]. Therefore, development of new therapeutic agents with novel hypoglycemic mechanism and fewer side effects is highly desired for the treatment of T2DM.

Human sodium-dependent glucose co-transporters (hSGLTs) are a group of membrane proteins responsible for glucose reabsorption from the glomerular filtrate in the proximal convoluted tubule. The mechanism by which hSGLTs mediate glucose reabsorption in the kidney is driven by the electrochemical sodium gradient between the glomerular filtrate and the proximal convoluted tubule [8]. Among them, hSGLT2, a low-affinity and high-capacity transporter expressed exclusively in the top third segment of the proximal convoluted tubule of the kidney, is responsible for 90% of renal glucose reabsorption [9] [10]. Inhibition of hSGLT2 induces glucose excretion from proximal convoluted tubule into urine and thereby reduces blood glucose concentration. Besides, hSGLT2 inhibitors have been confirmed to reduce body weight and prevent hypoglycemia [11]. Since hSGLT2 is mainly distributed in the kidney, hSGLT2 inhibitors would not cause common discomforts of other T2DM therapeutics agents, such as gastrointestinal flatulence, diarrhea and edema [12]. Moreover, recent studies have shown that hSGLT2 inhibitors also exhibit the capability to prevent patients from serious complications, including heart failure, nephropathy and liver

disease [13] [14] [15]. Therefore, hSGLT2 inhibitors have been considered as effective therapeutic agents with less side effects, thus can be served as monotherapy or combined with other hypoglycemic agents due to their distinctive hypoglycemic mechanism.

Phlorizin, an *O*-glucoside natural product substance, is known to be the first hSGLT2 inhibitor [16]. Although phlorizin has sufficient efficacy against hSGLT2, it is ultimately considered to be inappropriate as a hypoglycemic agent due to gastrointestinal side effects, low oral bioavailability and insufficient selectivity for hSGLT2. Subsequently, several *O*-glycoside inhibitors were developed to increase the selectivity for hSGLT2 but the shortcomings were not significantly improved [17]. Due to the fact that *C*-glycoside inhibitors are more metabolically stable because of their resistance to gastrointestinal  $\beta$ -glucosidases, they exhibit better efficacy and have fewer side effects comparing to *O*-glycoside inhibitors [18] [19] [20]. So far, three *C*-glycoside inhibitors, canagliflozin, dapagliflozin and empagliflozin, have been approved by the U.S. Food and Drug Administration (FDA) for the treatment of T2DM [21]. Although *C*-glycoside inhibitors exhibit sufficient efficacy, previous studies indicated that *N*-glycoside inhibitors may exhibit better potential to serve as a drug target for the treatment of T2DM due to their good metabolic stability, oral bioavailability, and low clearance [22] [23]. Comparing to glycoside inhibitors, non-glycoside inhibitors show several advantages, such as better tissue permeability, stability, serum half-time, less expensive and easier to synthesize [24] [25]. Therefore, development of novel *N*-glycoside and non-glycoside hSGLT2 inhibitors would provide greater advance in antidiabetic therapy.

Over the past few years, since the crystal structure of hSGLT2 is not yet available, the strategies for developing new hSGLT2 inhibitors are limited to chemical synthesis, structure-activity relationships (SAR) analysis, and lead modification [26] [27] [28]. These strategies would be high cost and time-consuming in the process of new drug development and the results are also limited to the highly structural similarity. Thus, in this study, we aimed to discover novel *N*-glycoside and non-glycoside hSGLT2 inhibitors through a combination of several computational approaches, such as ligand-based pharmacophore modeling, virtual screening, homology modeling, molecular docking, ADMET predictions, molecular dynamics (MD) simulations and binding free energy calculations. These computational approaches have been successfully and intensively used for the rapid assessment of small molecular databases in order to lower the cost and speed up the early stage of the new drug development process. Finally, one *N*-glycoside (NSC679207) and one non-glycoside (TCM\_Piperenol\_A) hSGLT2 inhibitors successfully identified from National Cancer Institute (NCI) and Traditional Chinese Medicine (TCM) databases, respectively, were shown to exhibit better binding affinities, pharmacokinetic properties and less toxicity comparing to the currently commercially available canagliflozin and can be further examined by a series of *in vitro* and/or *in vivo* bioassays to ensure their efficacy. Besides, our results also reveal several essential interactions between these two

compounds and the binding site of hSGLT2, which may ultimately contribute to the design of novel and potent hSGLT2 inhibitors for the clinical use in the future.

## 2. Materials and Methods

### 2.1. Preparation of the Training and Test Set Compounds

A total of 45 known hSGLT2 inhibitors were collected from the previous literatures [29]-[36]. Among them, 22 compounds, whose IC<sub>50</sub> values range from 0.3 to 4600 nM, were selected as the training set (**Figure 1(A)**); while the remaining 23 compounds, whose IC<sub>50</sub> values range from 2.39 to 1376 nM, were selected as the test set (**Figure 1(B)**). The 2D chemical structures of these compounds were sketched by ChemSketch and then converted into 3D structures by BIOVIA Discovery Studio (DS) 2017 R2. These compounds were subsequently optimized under the ionization state at pH 7.0 using the Prepare Ligand protocol of DS 2017 R2.

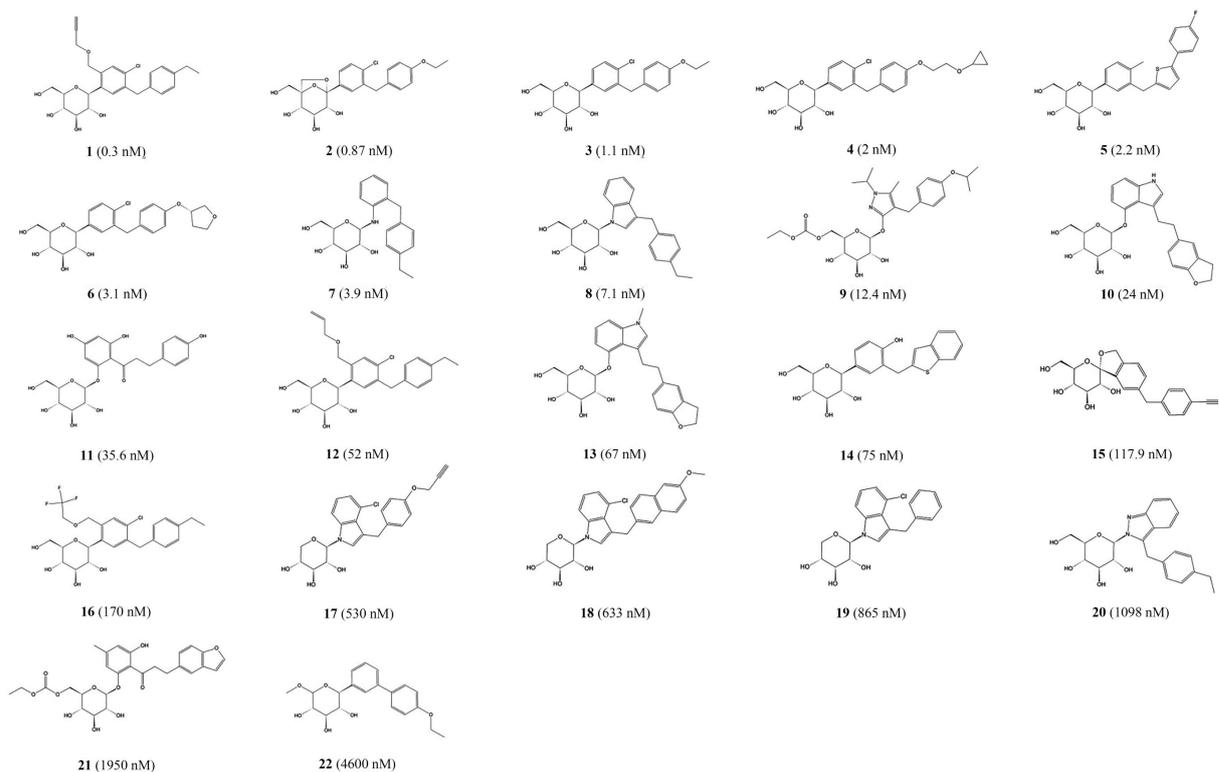
### 2.2. Ligand-Based Pharmacophore Model Generation

Before starting the pharmacophore model generation process, the conformers of each training set compound were generated by the BEST conformation model generation method and limited to a maximum conformer value of 255 with a 20 kcal/mol energy cut off. The multiple acceptable 3D conformations were obtained and then were inputted into DS 2017 R2 to generate 10 possible pharmacophore hypotheses using HypoGen module, which include hydrogen bond acceptor (HBA), hydrogen bond donor (HBD), hydrophobic (HY), ring aromatic (RA) and excluded volumes (EV). The chemical features were set to a minimum of 1 to a maximum of 5 and the maximum of EV was set to 5, while the other parameters were set to default values [37]. These generated hypotheses were ranked according to their total cost, correlation coefficient, and root mean square deviation (RMSD). In order to discriminate the quality of these hypotheses, two theoretical cost values, null cost and fixed cost, were evaluated. The null cost represents the highest cost of the model with no correlation, while the fixed cost represents the cost of the simplest model which fits all data perfectly. According to the above mentioned criteria, the hypothesis with the lowest total cost, the lowest RMSD value, and the highest correlation coefficient was considered as the best pharmacophore model.

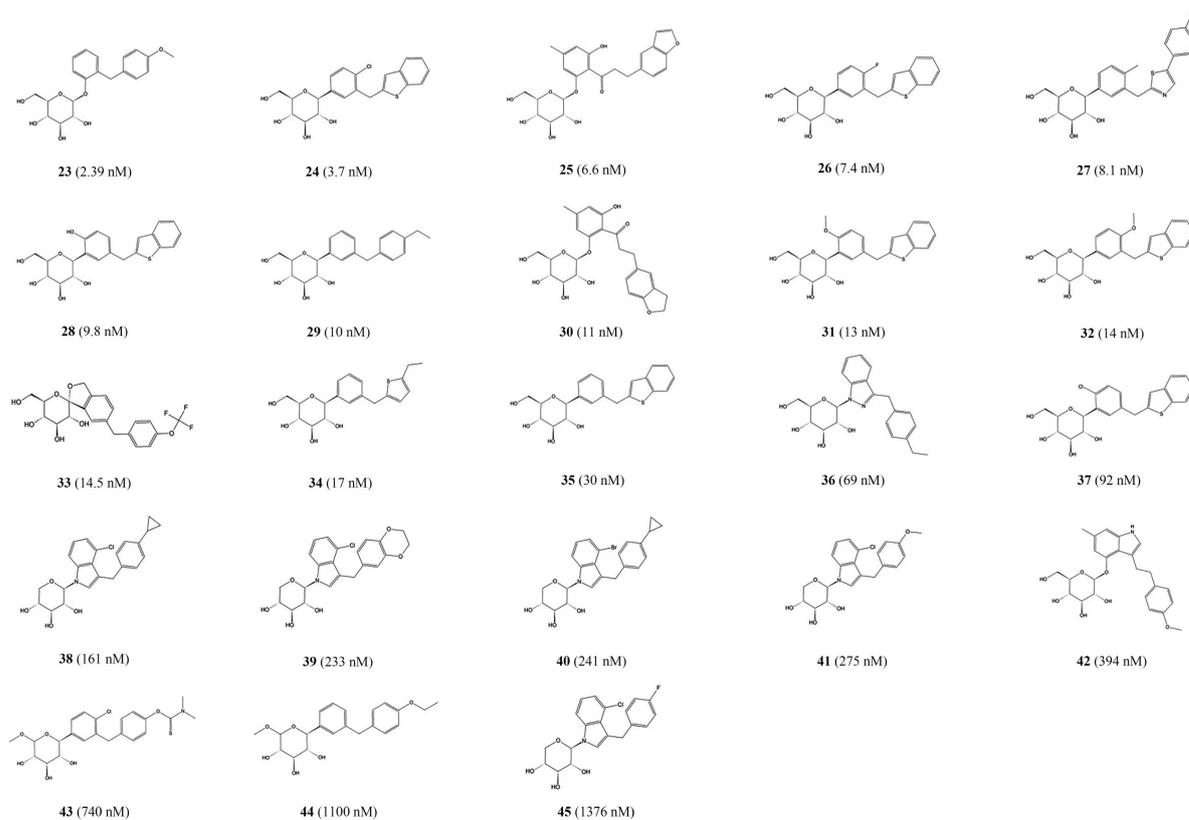
### 2.3. Ligand-Based Pharmacophore Model Validation

The quality of the best pharmacophore model was validated by test set prediction and Güner-Henry (GH) score method [38]. 23 test set compounds were used to elucidate whether the generated pharmacophore model is proficient to predict the activities of the compounds other than the training set compounds. GH score method was used to determine how well the generated pharmacophore model could differentiate potential hSGLT2 inhibitors from other non hSGLT2

(A)



(B)

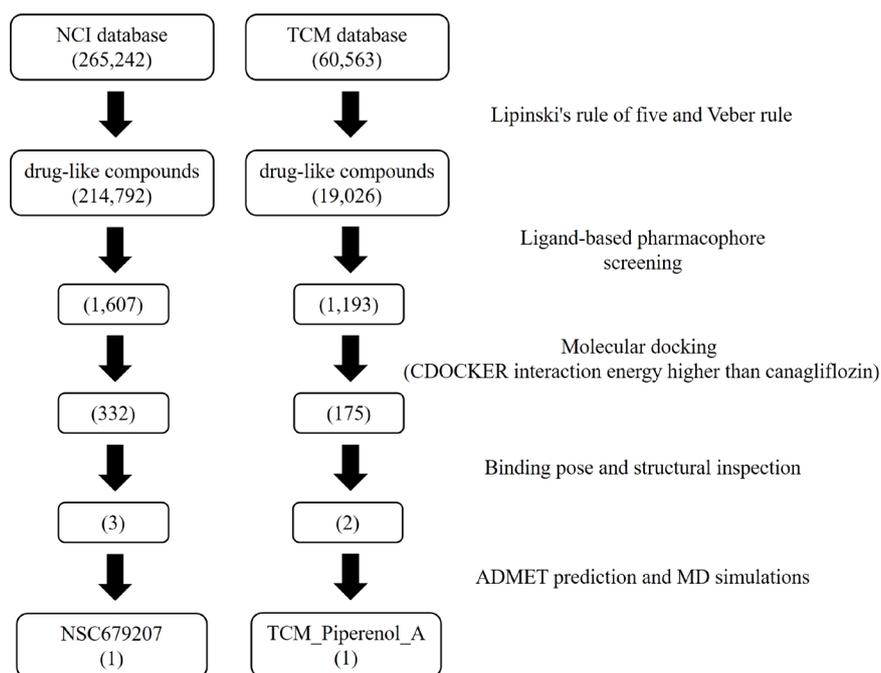


**Figure 1.** Chemical structures of 45 known hSGLT2 inhibitors used in training set and test set along with their experimental activity ( $IC_{50}$ ) values. (A) Compounds 1 - 22 serve as the training set and (B) compounds 23 - 45 compounds serve as the test set.

inhibitors. A total of 550 compounds, including 50 active compounds selected from the previous literatures [29]-[36] and 500 decoys from DUD-E website, were used to calculate the GH score. These 50 active compounds were different from the 45 active compounds in the training and test sets. All compounds were used to construct the 3D database (maximum of 255 conformations for each compound) using Build 3D Database of DS 2017 R2. Then, the generated pharmacophore model was employed for screening the 3D database using Screen Library with default parameters. A set of statistical parameters were obtained and then the GH score was calculated to evaluate the quality of the generated pharmacophore model.

## 2.4. Virtual Screening

The overall flowchart showing the entire virtual screening process and the screening criteria in this study are presented in **Figure 2**. First, the NCI and TCM databases were filtered by the Lipinski rule of five [39] and Veber rule [40] of DS 2017 R2 to select the compounds with good drug-like properties. Then, the well validated pharmacophore model was utilized as a 3D query to identify hSGLT2 inhibitors using Screen Library with default parameters. Compounds that successfully mapped all the features of the pharmacophore model were considered as potential hit compounds, which were further submitted to molecular docking. The hit compounds with higher binding affinities comparing to that of canagliflozin, the commercially available hSGLT2 inhibitor, were considered as novel hSGLT2 inhibitors. Finally, the pharmacokinetic properties and binding stabilities



**Figure 2.** The overall flowchart depicting each stage of the virtual screening process and the corresponding screening criteria in this study. The numbers shown in the parentheses indicate the number of compounds remained after each screening stage.

of these novel hSGLT2 inhibitors were further evaluated through ADMET prediction and MD simulations, respectively.

### 2.5. Homology Modeling

The amino acid sequence of hSGLT2 (accession code: NP\_003032) was obtained from the National Center for Biotechnology Information (NCBI) protein sequence data bank. Since the crystal structure of vSGLT (PDB ID: 2XQ2) [41], isolated from *Vibrio parahaemolyticus*, has the highest identities and similarities with hSGLT2 sequences, it was selected as the template to build the homology model of hSGLT2 using Build Homology Models of DS 2017 R2. Subsequently, the homology model obtained after 10 ns of MD refinement was evaluated via Ramachandran plot [42] on RAMPAGE website and Verify Protein (Profiles-3D) protocol of DS 2017 R2 to assess the compatibility of the 3D structure with the corresponding amino acid sequence [43].

### 2.6. Molecular Docking

Molecular docking was used to evaluate the binding affinity and clarify the binding mode of the screened compounds. Since vSGLT (PDB ID: 3DH4) [44] is the only co-crystallized structure of vSGLT and galactose, the position of galactose was used to identify the binding site of hSGLT2, with a sphere radius of 10 Å around the binding site being defined as the docking region [45]. CDOCKER program [46] of DS 2017 R2 with the CHARMM force field was used for precise and flexible docking. All of the docking parameters were set to default. For each screened compound, 10 docking poses were generated and the pose with the lowest CDOCKER interaction energy was compared to that of canagliflozin. The compounds with lower CDOCKER interaction energy than that of canagliflozin were finally selected.

### 2.7. ADMET Prediction

In this study, the online server admetSAR was employed to predict the ADMET (absorption, distribution, metabolism, excretion, and toxicity) properties of the screened compounds. It contains 22 qualitative classification and 5 quantitative regression models with highly predictive accuracy, allowing to estimate ADMET properties of the inputted compounds [47]. Each selected compound was converted to the simplified molecular-input line-entry systems (SMILES) format and inputted into the admetSAR for prediction. The compounds with superior ADMET properties were considered as potential hSGLT2 inhibitors.

### 2.8. Molecular Dynamics (MD) Simulation

The binding stabilities between hSGLT2 structure and the selected compounds were evaluated by MD simulations using Gromacs 2016.3 [48]. Combined with NVIDIA CUDA 8.0, Antechamber in AMBER tool 2017 and ACPYPE, the AMBER99SB-ILDN force field was applied for the hSGLT2 and the general

AMBER force field (GAFF) was applied for the screened compounds in the MD simulations process [49] [50]. All systems were immersed in a 10.0 Å dodecahedron periodic box with TIP3P water molecules. The sodium and chloride ions were added to reach 0.15 M as physiology state. The temperature was heated gradually from 0 to 310 K and then equilibrated to 310 K at 1.0 atm pressure. The entire system was stabilized through energy minimization and then executed two steps of restrained equilibrations for 200 ps, including NVT equilibration and NPT equilibration. Subsequently, the full MD simulations were performed for 150 ns in the equilibrate state.

## 2.9. Binding Free Energy Calculation

The Poisson-Boltzmann and surface area continuum solvation (MM/PBSA) method implemented in `g_mmpbsa` for GROMACS-5.1.X was used to calculate the binding free energies ( $\Delta G_{\text{binding}}$ ) between hSGLT2 and the selected compounds [51].  $\Delta G_{\text{binding}}$  was composed of four individual components, including van der Waals ( $\Delta G_{\text{vdw}}$ ), electrostatic ( $\Delta G_{\text{elec}}$ ), polar ( $\Delta G_{\text{polar}}$ ) and nonpolar ( $\Delta G_{\text{nonpolar}}$ ) solvation energy. The solvent accessible volume only (SAV-only) nonpolar model was applied for nonpolar solvation energy calculation in MM/PBSA. The temperature was set to 310 K, while the other parameters were set to default. For a more precise binding free energy analysis, the stable last 20 ns in the MD simulation trajectory was selected to generate a total of 2000 snapshots.

## 3. Results and Discussion

In the early stage of hSGLT2 inhibitor development, the strategies were committed to chemical synthesis, structure-activity relationships (SAR) analysis, and lead modification. However, no significant improvements in efficacy have been made due to the highly structure similarity. Through the aid of several computational approaches, precise molecular level study of hSGLT2-inhibitors binding interaction has become possible. To the best of our knowledge, this study is the first attempt to discover novel *N*-glycoside and non-glycoside hSGLT2 inhibitors through a combination of several computational approaches, including ligand-based pharmacophore modeling, virtual screening, homology modeling, molecular docking, ADMET predictions, molecular dynamics (MD) simulations and binding free energy calculations.

### 3.1. Pharmacophore Model Generation

Pharmacophore modeling has been proven to be very successful not only in demonstrating the structure-activity relationships between ligands and receptors, but also in developing new drugs. A pharmacophore model can be established either in a ligand-based manner or in a structure-based manner. While structure-based pharmacophore model works directly with the 3D structure of the target or its complex with a ligand, ligand-based pharmacophore model is derived from the structure of a series of ligands by analyzing the common pivotal

features of the superimposed ligands and translated into pharmacophore features subsequently. The generated pharmacophore model offers a rational hypothetical view of these significant chemical features responsible for activity. Since the crystal structure of hSGLT2 is not yet available, direct structure-based pharmacophore modeling is not applicable.

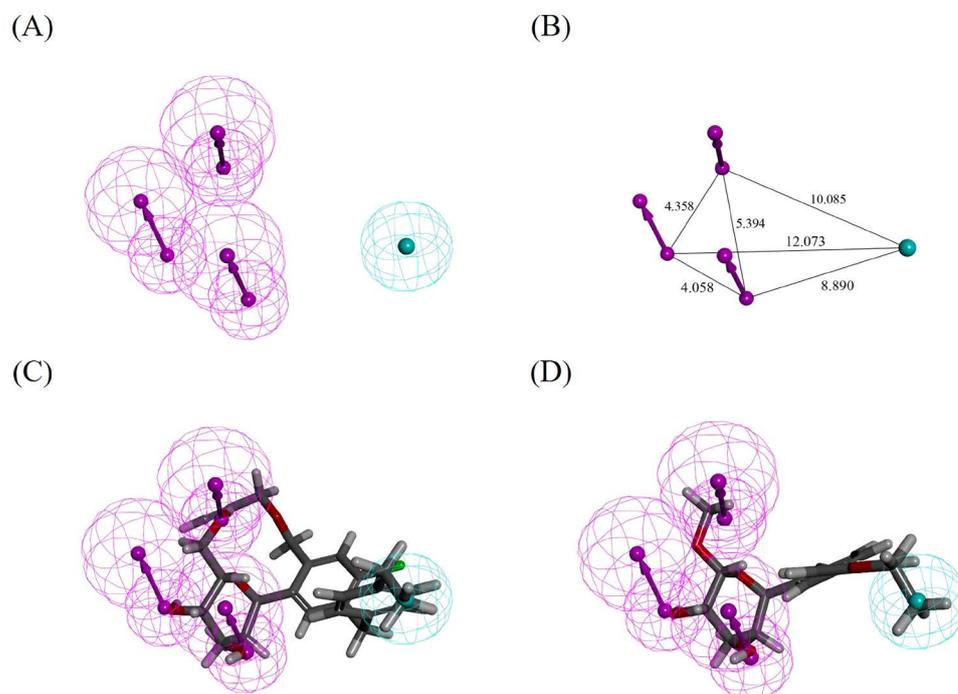
In this study, a total of 10 pharmacophore hypotheses were established based on the training set compounds (**Figure 1**) using the HypoGen module embedded in DS 2017 R2 and the results are given in **Table 1**. To choose the best hypothesis, a cost analysis was assessed on the basis of two important theoretical cost calculations (represented in bit units): the fixed cost, which represents the simplest model that fits all data perfectly, and the null cost, which represents the highest cost of a model with no correlation. The total cost of an excellent pharmacophore model should be close to the fixed cost and far away from the null cost. The RMSD value shows the quality of the correlation between the experimental and the estimated activity data. To verify the predictive ability of all hypotheses on the training set compounds, the activity of each compound was estimated by the regression analysis [52]. According to the above mentioned criteria, the first hypothesis (Hypo1) was chosen to be the best hypothesis characterized by the lowest total cost (110.381), the highest correlation coefficient ( $R = 0.919$ ) and the lowest RMSD value (0.993).

**Figure 3(A)** and **Figure 3(B)** show the 3D spatial arrangement and distance constraints of the features (three HBDs and one HY) in Hypo1. The features of Hypo1 were mapped onto the most active compound 1 with a fit value of 10.616 and the least active compound 22 with a fit value of 7.452 as shown in **Figure 3(C)** and **Figure 3(D)**, respectively. Fit value indicates how well the features in a pharmacophore model overlap the chemical features in the molecule and thus

**Table 1.** The parameters of the top ten HypoGen pharmacophore hypotheses generated by the 22 training set compounds using HypoGen module of DS 2017 R2.

Hypo No.	Total cost	Cost difference <sup>a</sup>	RMSD	Correlation	Features <sup>b</sup>	Max. fit
Hypo 1	110.381	30.556	0.993	0.919	HBD, HBD, HBD, HY, 4EV	10.616
Hypo 2	113.815	27.122	1.145	0.890	HBD, HBD, HBD, HY, 2EV	10.526
Hypo 3	113.934	27.003	1.157	0.887	HBA, HBD, HBD, HY, 2EV	10.416
Hypo 4	114.804	26.133	1.136	0.895	HBA, HBD, HBD, HY, 3EV	11.161
Hypo 5	116.672	24.265	1.238	0.870	HBA, HBD, HY, RA, 3EV	10.759
Hypo 6	117.635	23.302	1.271	0.863	HBA, HBD, HY, RA, EV	10.779
Hypo 7	117.675	23.262	1.164	0.898	HBA, HBD, HBD, HY, 4EV	12.029
Hypo 8	117.981	22.956	1.279	0.862	HBD, HBD, HY, RA, 3EV	10.854
Hypo 9	120.500	20.437	1.393	0.830	HBD, HBD, HBD, HY, EV	10.390
Hypo 10	121.092	19.845	1.437	0.815	HBA, HBA, HBD, HY	9.865

<sup>a</sup>Cost difference = null cost – total cost. null cost = 140.937. Fixed cost = 97.198. All cost values are in bits; <sup>b</sup>HBA, hydrogen bond acceptor; HBD, hydrogen bond donor; HY, hydrophobic; RA, ring aromatic; EV, excluded volumes.



**Figure 3.** The best pharmacophore model (Hypo 1) of hSGLT2 inhibitors generated by HypoGen module of DS 2017 R2. (A) There are 3HBDs and 1HY features for Hypo 1. (B) 3D spatial arrangement and geometric parameters of Hypo 1. (C) The most active compound 1 (0.3 nM) maps onto Hypo 1. (D) The least active compound 22 (4600 nM) maps onto Hypo 1. The features of HBD and HY are shown in magenta and cyan, respectively.

aid in understanding the chemical meaning of the hypothesis. The difference between the fit values of these two compounds may explain the difference in their activities.

### 3.2. Pharmacophore Model Validations

The accuracy of Hypo1 is important to achieve success in the subsequent virtual screening process, therefore its validity was assessed by test set prediction and GH score method.

#### 3.2.1. Test Set Prediction

The significance of the selected pharmacophore model depends on its ability to predict the biological activity of test set compounds along with the training set compounds. A total of 23 structurally diverse test set compounds with a wide range of activities were utilized to verify the prediction ability of Hypo1. All the test compounds were subjected to the same conformation generation procedures as for the training set compounds. **Table 2** shows the experimental and estimated activities of the 22 training set compounds and **Table 3** shows the experimental and estimated activities of the 23 test set compounds.

To verify the predictability power of Hypo1, the estimated activity ( $pIC_{50}$ ) for each individual test set compound was calculated in order to correlate to its experimental activity using simple regression analysis. The obtained correlation

**Table 2.** Experimental and estimated IC<sub>50</sub> values of the training set compounds.

Compound. No.	Exp. IC <sub>50</sub> (nM)	Est. IC <sub>50</sub> (nM)	Error <sup>a</sup>	Fit value <sup>b</sup>	Exp. scale <sup>c</sup>	Est. scale <sup>c</sup>
1	0.3	0.74	+2.5	10.616	+++	+++
2	0.87	6.77	+7.8	9.322	+++	+++
3	1.1	2.26	+2.1	9.798	+++	+++
4	2	2.12	+1.1	9.825	+++	+++
5	2.2	6.41	+2.9	9.346	+++	+++
6	3.1	1.17	-2.6	10.084	+++	+++
7	3.9	4.68	+1.2	9.483	+++	+++
8	7.1	3.37	-2.1	9.625	+++	+++
9	12.4	47.84	+3.9	8.473	++	++
10	24	6.57	-3.7	9.335	++	+++
11	35.6	198.24	+5.6	7.856	++	++
12	52	28.95	-1.8	8.691	++	++
13	67	29.19	-2.3	8.688	++	++
14	75	99.22	+1.3	8.156	++	++
15	117.9	346.19	+2.9	7.613	++	++
16	170	267.46	+1.6	7.725	++	++
17	530	237.36	-2.2	7.777	++	++
18	633	338.49	-1.9	7.623	++	++
19	865	845.51	-1.0	7.226	++	++
20	1098	318.73	-3.4	7.649	+	++
21	1950	636.22	-3.1	7.349	+	++
22	4600	502.24	-9.2	7.452	+	++

<sup>a</sup>Error factor calculated as the ratio of the experimental activity to the estimated activity; “+” indicates that the estimated IC<sub>50</sub> is higher than the experimental IC<sub>50</sub>; “-” indicates that the estimated IC<sub>50</sub> is lower than the experimental IC<sub>50</sub>; <sup>b</sup>Fit value indicates how well the compound mapped onto the features of Hypo 1; <sup>c</sup>Activity scale: highly active (IC<sub>50</sub> < 10 nM, +++), moderately active (10 nM ≤ IC<sub>50</sub> < 1000 nM, ++), and inactive (IC<sub>50</sub> ≥ 1000 nM, +).

**Table 3.** Experimental and estimated IC<sub>50</sub> values of the test set compounds.

Compound. No.	Exp. IC <sub>50</sub> (nM)	Est. IC <sub>50</sub> (nM)	Error <sup>a</sup>	Fit value <sup>b</sup>	Exp. scale <sup>c</sup>	Est. scale <sup>c</sup>
23	2.39	3.42	+1.4	9.619	+++	+++
24	3.7	5.00	+1.4	9.453	+++	+++
25	6.6	4.32	-1.5	9.518	+++	+++
26	7.4	3.42	-2.2	9.619	+++	+++
27	8.1	6.07	-1.3	9.370	+++	+++
28	9.8	6.52	-1.5	9.339	+++	+++
29	10	8.51	-1.2	9.223	++	+++
30	11	3.94	-2.8	9.557	++	+++

## Continued

31	13	18.64	+1.4	8.882	++	++
32	14	23.28	+1.7	8.786	++	++
33	14.5	5.92	-2.5	9.380	++	+++
34	17	14.93	-1.1	8.979	++	++
35	30	19.37	-1.5	8.865	++	++
36	69	44.39	-1.6	8.505	++	++
37	92	34.19	-2.7	8.619	++	++
38	161	47.11	-3.4	8.480	++	++
39	233	186.73	-1.2	7.881	++	++
40	241	64.45	-3.8	8.350	++	++
41	275	133.14	-2.1	8.028	++	++
42	394	262.75	-1.5	7.733	++	++
43	740	798.61	+1.1	7.250	++	++
44	1100	379.60	-2.9	7.573	+	++
45	1376	669.61	-2.1	7.327	+	++

<sup>a</sup>Error factor calculated as the ratio of the experimental activity to the estimated activity; “+” indicates that the estimated IC<sub>50</sub> is higher than the experimental IC<sub>50</sub>; “-” indicates that the estimated IC<sub>50</sub> is lower than the experimental IC<sub>50</sub>; <sup>b</sup>Fit value indicates how well the compound mapped onto the features of Hypo 1; <sup>c</sup>Activity scale: highly active (IC<sub>50</sub> < 10 nM, +++), moderately active (10 nM ≤ IC<sub>50</sub> < 1000 nM, ++), and inactive (IC<sub>50</sub> ≥ 1000 nM, +).

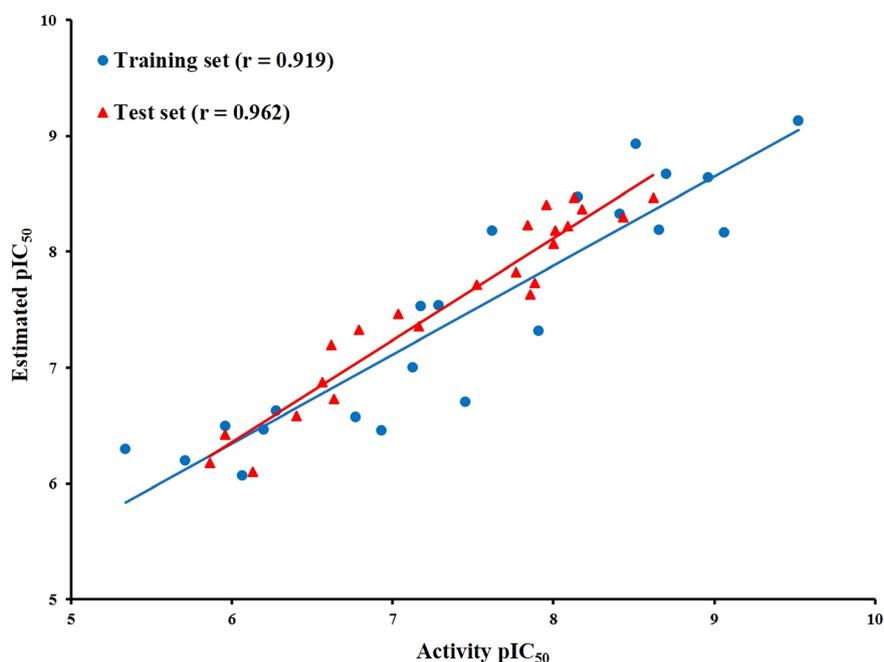
coefficient values were 0.962 and 0.919 for the test set and training set compounds, respectively (Figure 4). The results demonstrate that Hypo1 is a reliable pharmacophore model with high predictability power.

### 3.2.2. GH Score Method

The main purpose of the GH score method is to assess the discriminative ability of Hypo1 to correctly identify the known active compounds from the inactive ones. All the detailed GH score parameters of Hypo1 are summarized in Table 4. To perform GH score calculation, an external database, consisting of a total of 550 compounds (D) including 50 actives (A), was utilized to evaluate the discriminative ability of Hypo1. After screening this database by Hypo1, 57 compounds were retrieved as hits. Among them, 49 compounds were from the known actives (Ha). According to the previous study, the GH score of a good screening model must be higher than 0.7 [53]. The GH score of Hypo1 was calculated to be 0.876, indicating that it has the ability in selecting the active compounds among inactive ones.

### 3.3. Virtual Screening

A schematic flowchart of the overall virtual screening process in this study is provided in Figure 2. To effectively screen the NCI and TCM databases with 265,242 and 60,563 compounds, respectively, Lipinski rule of five [39] and Veber



**Figure 4.** Plot of the correlation coefficient ( $r$ ) between the experimental and the estimated activity values ( $pIC_{50}$ ) based on Hypo 1 model for the 22 training set compounds (blue circle) and the 23 test set compounds (red triangle).

**Table 4.** Validation of Hypo 1 pharmacophore model by Güner-Henry (GH) score method.

Parameter	Values
Total number of compounds in database (D) <sup>a</sup>	550
Total number of actives in database (A)	50
Total number of hit compounds from the database (Ht) <sup>b</sup>	57
Total number of active compounds in hit list (Ha)	49
%Yield of actives (%Y) [(Ha/Ht) × 100]	85.96
%Ratio of actives (%R) [Ha/A] × 100]	98.00
False negatives [A – Ha]	1
False positives [Ht – Ha]	8
Goodness of hit (GH) <sup>c</sup>	0.876

<sup>a</sup>D is the number of total compounds in database, including 50 active compounds and 500 decoys; <sup>b</sup>Ht is the total number of hit compounds in database, including 49 active compounds and 8 decoys;

<sup>c</sup> $\left[ \left( \frac{Ha}{4Ht \times A} \right) (3A + Ht) \times \left( 1 - \left( \frac{Ht - Ha}{D - A} \right) \right) \right]$ , GH score > 0.7 indicates a good pharmacophore model.

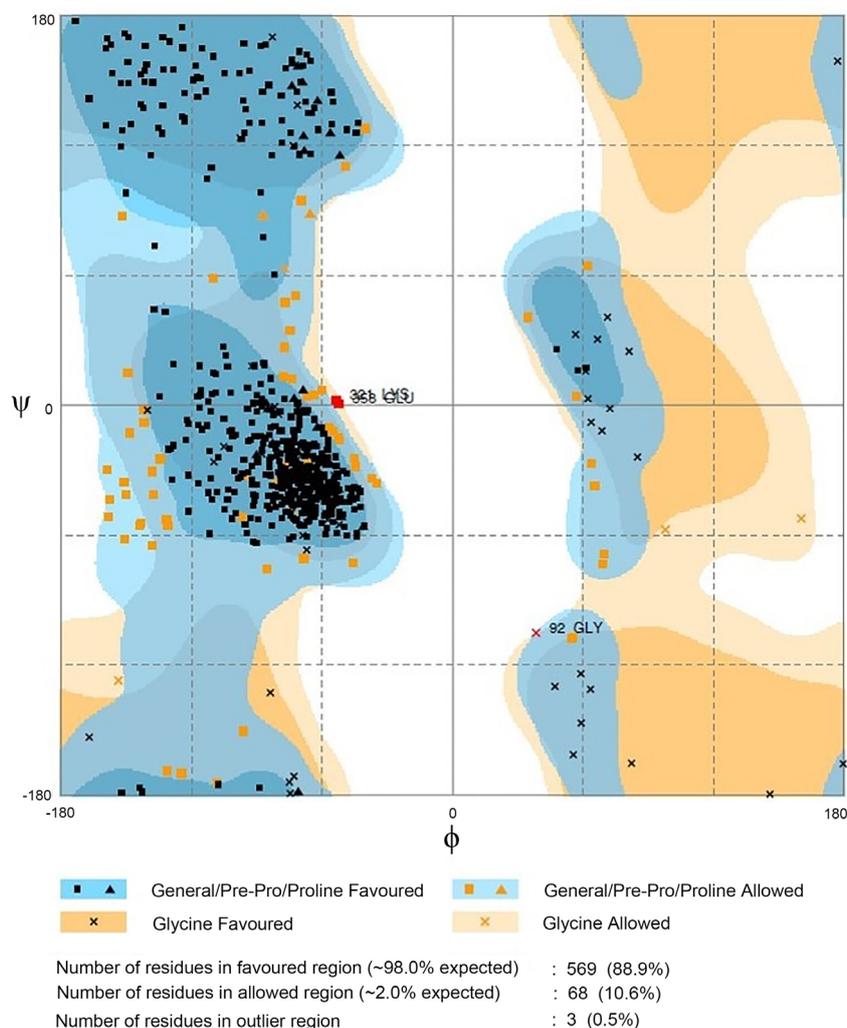
rule [40] were first employed to ensure that the screened compounds are drug-like. The remaining 214,792 and 19,026 compounds that passed these rules were subsequently mapped onto Hypo1 to identify potential hSGLT2 inhibitors. Finally, 1617 and 1193 compounds from NCI and TCM databases, respectively, which pass all the criteria mentioned above were subjected to further molecular docking studies. In this stage, our Hypo1 model shows great capability of efficiently identifying potential hits from a large number of compounds.

### 3.4. Homology Modeling

Homology modeling is an important method of generating the 3D structure of a target protein from a closely related template. The accuracy of the obtained structural model dramatically depends on the identity and similarity between the target and the template proteins. Our sequence alignment results indicate that the template vSGLT (PDB ID: 2XQ2) has a sequence identity of 24.0% and a similarity of 47.0% with hSGLT2. It is generally assumed that the sum of the BLAST sequence identity and similarity of at least 60% is suggestive of an effective sequence template. Thus, the crystal structure of vSGLT was herein adopted as the template and the pair-wise sequence alignment was used to construct the homology model of hSGLT2. The 3D structural model of hSGLT2, including 643 amino acids, was generated using Build Homology Models of DS 2017 R2. The model obtained after 10 ns of MD simulation refinement was further validated using a Ramachandran plot (RC) and the Verify Protein (Profiles-3D) protocol. The corresponding RC plot showed that 99.5% of the residues in the hSGLT2 structural model were found in both the most favored and additional allowed regions, and only 0.5% of the disfavored residues were found to be located in the loop part of the protein. It indicates that the torsion angles of the residues in the homology model are reliable [54]. The verification score obtained from Profile-3D upon inspecting the quality of the protein structure geometry was 193.37, close to the expected high verification score of 233.98, indicating that this homology model yields satisfactory results. We further examined that some structural features of this homology model were consistent with the characteristics of the SGLT family, which includes 14 transmembrane helices with inverted repeat topology segments (Figure 5).

### 3.5. Molecular Docking

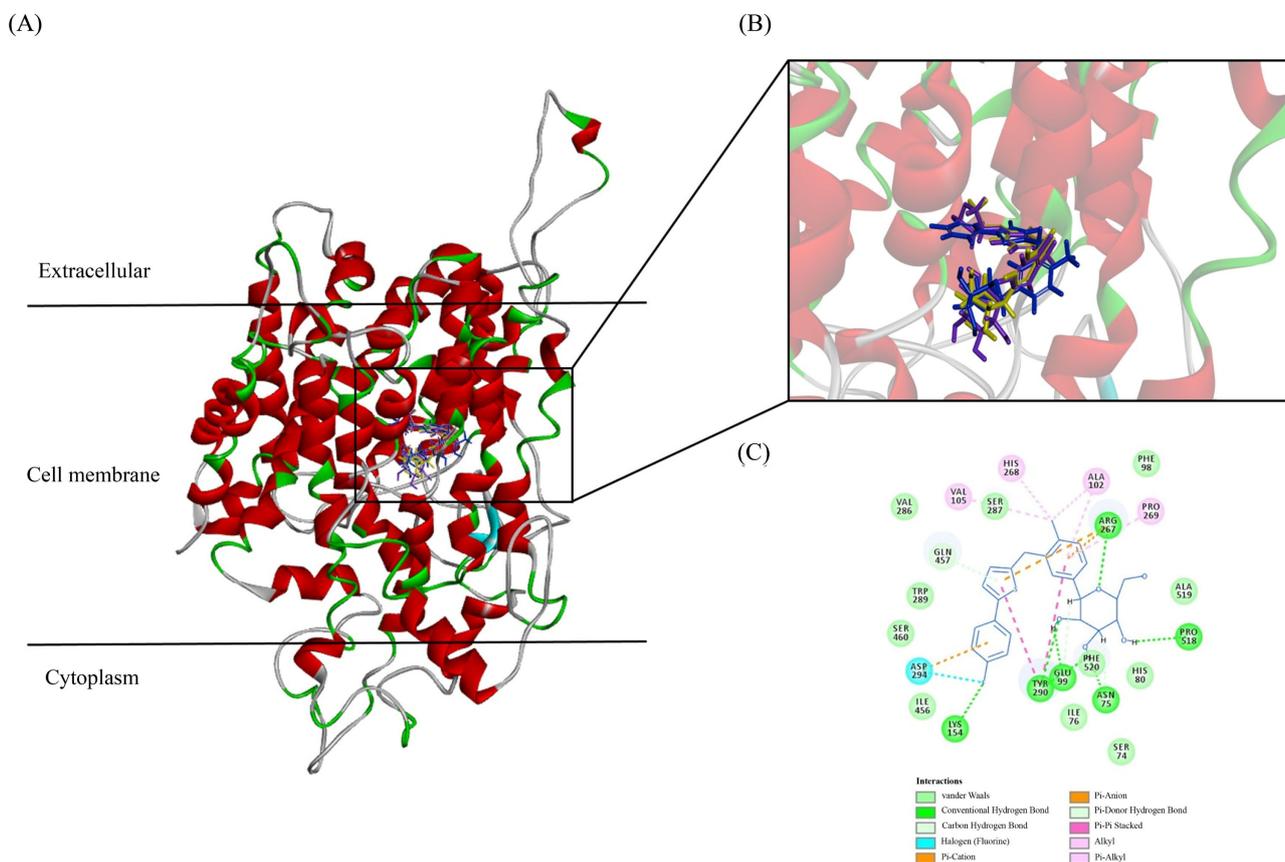
Molecular docking has been intensively used to clarify the binding mode of the compounds and to obtain the other information that could be utilized for further structural optimization. Although the structural model of hSGLT2 was built by homology modeling successfully, the binding site of hSGLT2 is still undefined. So far, vSGLT (PDB ID: 3DH4) [44] is the only co-crystallized structure of vSGLT and galactose. Therefore, the position of galactose was used to identify the binding site of hSGLT2 in this study. To distinguish the binding affinity of each screened compound, CDOCKER interaction energy (kcal/mol) was used as the selecting criteria. Three approved hSGLT2 inhibitors, including canagliflozin, dapagliflozin and empagliflozin, were used to determine the possible binding modes of the selected hSGLT2 inhibitor. These three hSGLT2 inhibitors show the identical binding mode in the binding site of hSGLT2 (Figure 6(A)). Among them, canagliflozin has the lowest CDOCKER interaction energy of  $-47.66$  kcal/mol (Figure 6(C)). Then, the 1607 and 1193 compounds selected from NCI and TCM databases, respectively, were docked into the binding site of hSGLT2 and the compounds with lower CDOCKER interaction energies than



**Figure 5.** Ramachandran plot of the hSGLT2 homology model.

that of canagliflozin were selected as potential hits, resulting in 332 and 175 compounds from NCI and TCM databases, respectively. These selected compounds were subjected to visual inspection to confirm their binding modes and whether their structural features belong to either *N*-glycoside or non-glycoside.

Previous studies have indicated that the hSGLT2 inhibitors fit into the binding site of hSGLT2 by interacting with several important residues, such as Asn75, His80, Glu99, Ala102, Arg267, Tyr290, Trp291, Gln457, and Pro518 [55] [56] [57] [58]. Among these important residues, Asn75, Glu99, Arg267, and Tyr290 are the most essential key residues since they form a hydrogen bonding network with hSGLT2 inhibitors [55] [56] [57] [58]. According to the above screening criteria, only 1 *N*-glycoside and 4 non-glycoside compounds were finally identified to be potential hSGLT2 inhibitors (Table 5). These screened compounds combined with essential key residues and formed stable combinations through hydrogen bonds. These selected compounds, which exhibit similar binding mode and higher binding affinities than canagliflozin, were further evaluated through ADMET prediction followed by MD simulation studies.

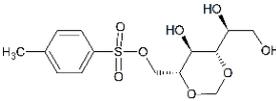
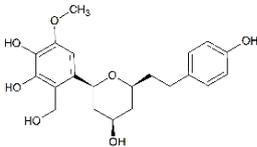
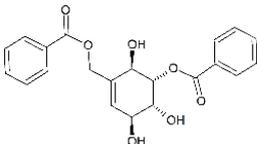


**Figure 6.** The binding mode and position of the three commercially available hSGLT2 inhibitors: (canagliflozin (blue), dapagliflozin (yellow) and empagliflozin (purple)). (A) Structure of the hSGLT2-inhibitors binding complex. (B) The binding mode of the three commercially available hSGLT2 inhibitors in the hSGLT2 binding site. (C) The 2D interaction diagram of canagliflozin with hSGLT2, showing the best CDOCKER interaction energy ( $-47.66$  kcal/mol) among these three inhibitors. The important residues involved in the hSGLT2 binding are labeled.

**Table 5.** The NSC and TCM number, 2D structure, category and CDOCKER interaction energy of the selected compounds.

Compound	2D structure	Category	CDOCKER interaction energy (kcal/mol)
Cnangliflozin		C-glycoside	$-47.66$
NSC679207		N-glycoside	$-48.33$
NSC727709		Non-glycoside	$-66.32$

## Continued

NSC231775		Non-glycoside	-55.81
TCM21610		Non-glycoside	-50.85
TCM_Poperenol_A		Non-glycoside	-49.63

## 3.6. ADMET Prediction

Prediction of the compound ADMET properties describes the disposition of a pharmaceutical compound within the human body. Considering that most of the previously reported hSGLT2 inhibitors failed in their clinical trials due to poor ADMET properties, filtering and optimization of the ADMET properties in the early stage of the drug development are necessary to avoid failure. In this study, the ADMET properties of canagliflozin and the selected 1 *N*-glycoside (NSC679207) and 4 non-glycoside (NSC727709, NSC231775, TCM21610 and TCM\_Piperenol\_A) compounds were predicted by the online sever admetSAR and the results are presented in **Table 6**. In terms of absorption, only NSC231775 have bad property in human intestinal absorption (HIA), suggesting oral delivery is infeasible. In terms of toxicity, canagliflozin, NSC679207, TCM21610 and TCM\_Piperenol\_A are all non-mutagenic (AMES toxicity test) and non-carcinogenic (carcinogenic profile) [59]. In human heart, hERG potassium channels are essential for regulating electrical activity. The blockage of hERG channel by drug causes lethal long QT syndrome (LQTS), which leads to withdraw from the clinical trials or markets [60]. The results of hERG inhibitor prediction I/II showed that TCM21610 is an hERG inhibitor, involving high LQTS risk than the others.

Since the US Food and Drug Administration (FDA) published the first *in vitro/in vivo* drug interaction guidance documents in 1997 and 1999, respectively, off-target transporter interactions and drug-drug interactions (DDIs) have become the major challenges for drug development. CYP450 enzymes-based unanticipated DDIs and drug metabolism problems are also a common cause of adverse drug events (ADE). Interacting with CYP450 enzymes may lead to DDIs, hepatotoxicity, reactive metabolite formation, idiosyncratic adverse drug interactions, and/or even loss the efficacy of drugs. CYP1A2 is related to alcohol metabolism, indicating that the inhibition of CYP1A2 results in the reducing of alcohol clearance and increases the risk of liver toxicity [61]. The inhibition of CYP2C19 will affect the metabolism of other drugs through the CYP2C19 pathway,

**Table 6.** ADMET properties of canagliflozin and the five selected compounds predicted using admetSAR online server. The gray shades indicate unfavorable properties.

Property	Compound	Canagliflozin	NSC679207	NSC727709	NSC231775	TCM21610	TCM_Piperenol_A
Human Intestinal Absorption <sup>a</sup>		+	+	+	-	+	+
P-glycoprotein Substrate <sup>b</sup>		+	-	-	+	+	+
P-glycoprotein Inhibitor I <sup>c</sup>		-	-	-	+	-	-
P-glycoprotein Inhibitor II <sup>c</sup>		-	-	-	-	+	-
Renal Organic Cation Transporter <sup>c</sup>		-	-	-	-	-	-
Blood-Brain Barrier <sup>d</sup>		+	+	+	+	-	-
CYP450 2C9 Substrate <sup>b</sup>		-	-	-	-	-	-
CYP450 2D6 Substrate <sup>b</sup>		-	-	-	-	-	-
CYP450 3A4 Substrate <sup>b</sup>		-	+	-	-	-	-
CYP450 1A2 Inhibitor <sup>c</sup>		-	-	+	-	-	-
CYP450 2C9 Inhibitor <sup>c</sup>		-	-	-	-	-	-
CYP450 2D6 Inhibitor <sup>c</sup>		-	-	-	-	-	-
CYP450 2C19 Inhibitor <sup>c</sup>		-	-	+	-	-	-
CYP450 3A4 Inhibitor <sup>c</sup>		-	-	-	-	-	-
CYP450 Inhibitory Promiscuity		high	low	low	low	low	low
hERG inhibition I <sup>c</sup>		-	-	-	-	-	-
hERG inhibition II <sup>c</sup>		-	-	-	-	+	-
Ames test <sup>e</sup>		-	-	+	-	-	-
Carcinogenicity <sup>f</sup>		-	-	-	+	-	-

<sup>a</sup>+ for positive and - for negative; <sup>b</sup>+ for substrate and - for non-substrate; <sup>c</sup>+ for inhibitor and - for non-inhibitor; <sup>d</sup>+ for penetrable and - for non-penetrable; <sup>e</sup>+ for toxic and - for non-toxic; <sup>f</sup>+ for carcinogens and - for non-carcinogens.

which is related to the function of platelet in human body. Therefore, CYP2C19 inhibitors should be carefully used while cooperating with drugs that affect platelet aggregation [62]. The level of CYP inhibitory promiscuity inhibition represents the risks of causing CYP450 enzymes-based unanticipated DDIs and metabolism problems [63]. According to the above information, canagliflozin and NSC727709 must pay attention to unanticipated DDIs and the risk of hepatotoxicity.

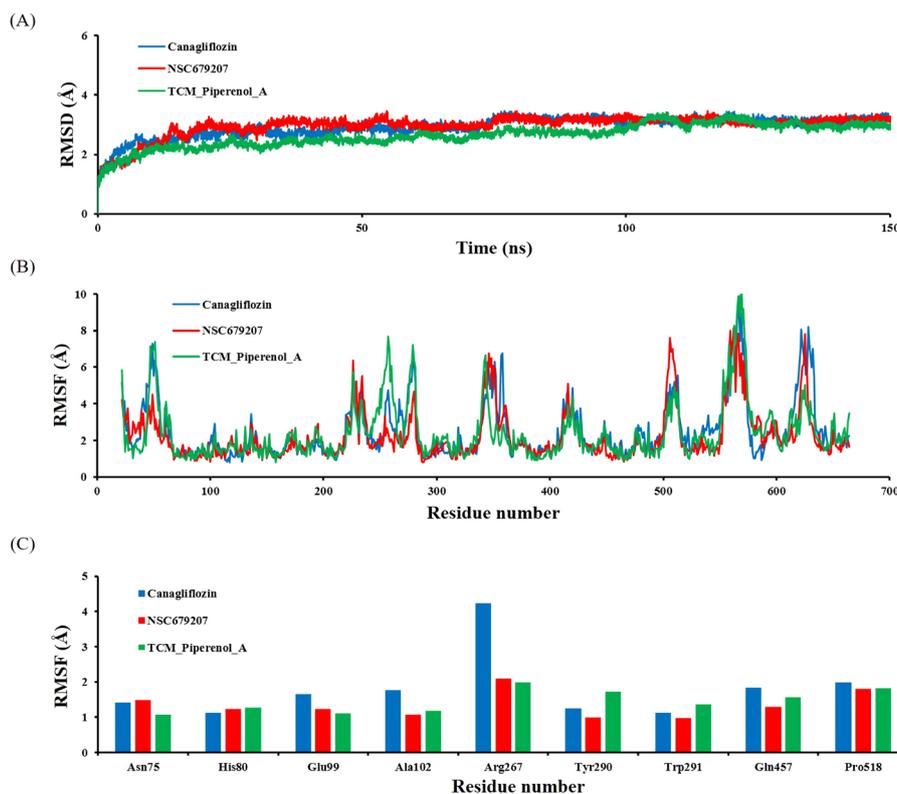
Based on the ADMET prediction, among the five selected compounds, only NSC679207 and TCM\_Piperenol\_A exhibit better ADMET properties than that of canagliflozin, indicating that both compounds exhibit great potential to serve as *N*-glycoside and non-glycoside hSGLT2 inhibitors.

### 3.7. MD Simulations

MD simulations are able to analyze the trajectory of protein-ligand complex during the dynamics process to validate the stability of docking results and calculate binding free energies in the entire process. The structures of hSGLT2 docked with canagliflozin, NSC679207, and TCM\_Piperenol\_A were further

subjected to 150 ns MD simulations to investigate the binding stabilities of these three compounds towards hSGLT2. The backbone RMSD of the three compounds and hSGLT2 revealed that an equilibrated and converged state was achieved after 50 ns of simulation. Therefore, MD simulations were then extended to 150 ns in an effort to gauge the stabilities of these three binding systems [64]. As shown in **Figure 7(A)**, no major fluctuations were observed. Canagliflozin, NSC679207, and TCM\_Piperenol\_A were maintained at stable state with averaged RMSD values of 3.195 Å, 3.178 Å, and 2.973 Å, respectively in the last 10 ns MD simulations. It allows the changes in the conformations of these three compounds and the residues of the binding site to be discerned.

To explore the protein residues flexibility throughout the MD simulation, the root-mean-squared fluctuations (RMSF) of individual hSGLT2 residues were calculated. As shown in **Figure 7(B)**, NSC679207 and TCM\_Piperenol\_A exhibited similar RMSF distributions comparing to canagliflozin. In addition, the conformational changes of the above mentioned hSGLT2 key residues (Asn75, Glu99, Arg267, and Tyr290) were all very small (**Figure 7(C)**). Our results show that both NSC679207 and TCM\_Piperenol\_A exhibit stable RMSF distribution, indicating that they can sufficiently fix most of the hSGLT2 key residues (**Figure 7(C)**). The results of RMSF analyses suggest that both NSC679207 and TCM\_Piperenol\_A have extremely stable binding stabilities, similar to that of canagliflozin, indicating



**Figure 7.** RMSD and RMSF analysis results during the 150 ns MD simulations. (A) RMSD values of the backbone atoms for hSGLT2-inhibitor complexes. (B) RMSF values of each hSGLT2 residue. (C) RMSF values of each hSGLT2 important residues.

that they form tight binding complexes with hSGLT2 during the entire MD simulation processes.

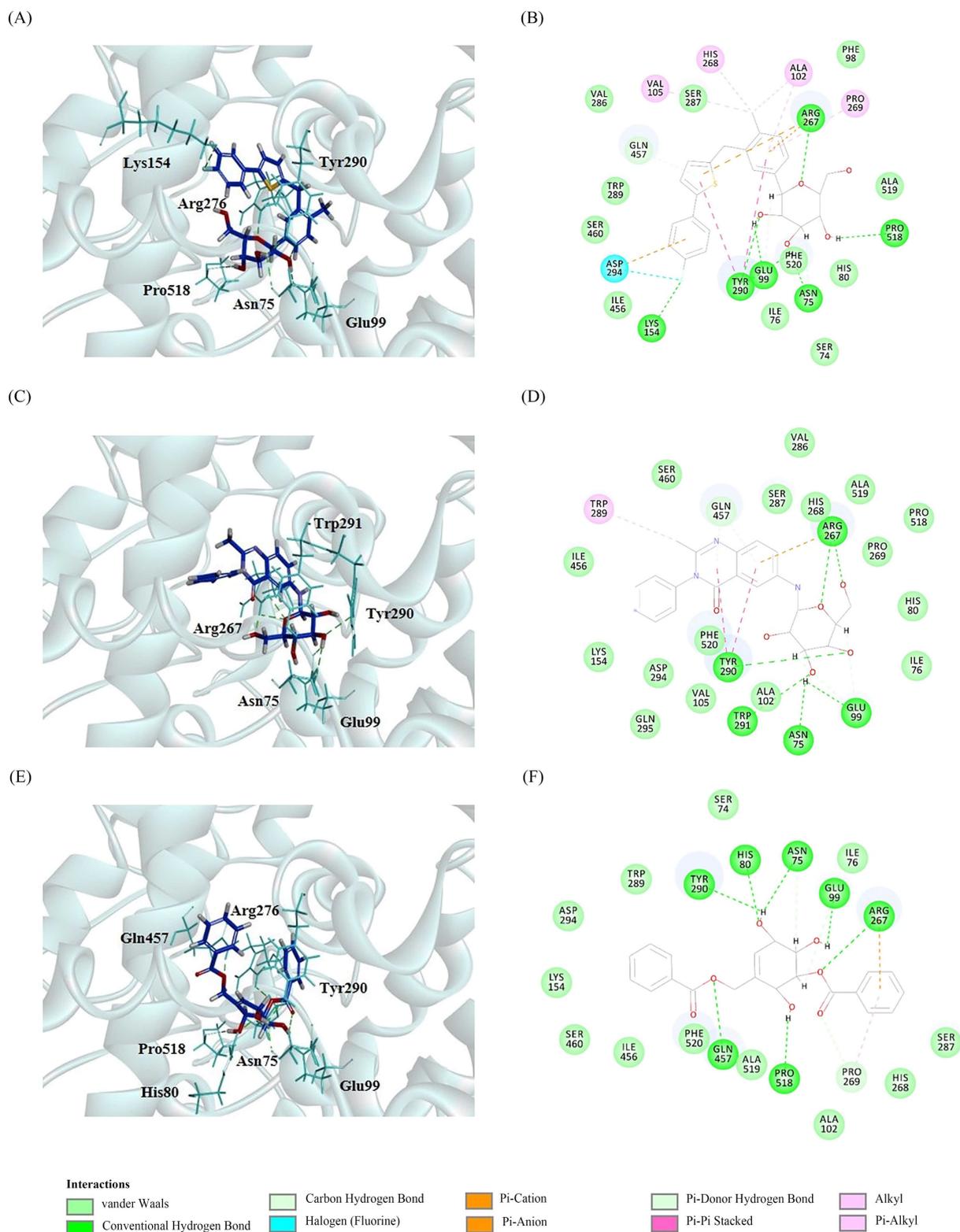
### 3.8. Binding Free Energy Calculation and the Analysis of Binding Interactions

In the past, molecular mechanics/Poisson-Boltzmann surface area (MM/PBSA) or molecular mechanics/generalized Born surface area (MM/GBSA) methods have been intensively applied to determine the correlations between the calculated free energy and the experimental activity. In this study, the binding free energy was calculated using the MM/PBSA method implemented in GROMACS-5.1.X 2000 snapshots from the last 20 ns of the MD simulations of the complexes of hSGLT2 with canagliflozin, NSC679207, and TCM\_Piperenol\_A were used to perform binding free energy ( $\Delta G_{\text{binding}}$ ) calculations and the results are as summarized in **Table 7**. NSC679207 (−45.30 kcal/mol) and TCM\_Piperenol\_A (−55.19 kcal/mol) showed higher and lower binding free energies, respectively, comparing to that of canagliflozin (−52.32 kcal/mol).

All glycoside hSGLT2 inhibitors contain a glycoside structure, which binds to the glucose binding site of hSGLT2 and an aglycone tail, which binds to the wall of the hydrophilic cavity leading to enhance the stability [65]. The most important binding affinities of the hSGLT2 inhibitors were provided by hydrogen bonding network around the glycoside structure. To understand the difference of the binding free energies among canagliflozin, NSC679207, and TCM\_Piperenol\_A, the initial binding interactions of these three compounds were revealed through their 3D structures and 2D binding interactions diagrams (**Figure 8**). According to **Figure 8(A)**, **Figure 8(C)**, and **Figure 8(E)**, these three compounds exhibit the identical binding mode and stably locate in the binding pocket of hSGLT2 through the above mentioned hydrogen bonding network. The 2D interactions (**Figure 8(B)**, **Figure 8(D)**, and **Figure 8(F)**) also showed that these three compounds are all able to form the essential hydrogen bonds with hSGLT2 through several key residues, including Asn75, Glu99, Arg267, and Tyr290. The difference of the binding free energies between canagliflozin and NSC679207 (*N*-glycoside compound) are mostly occurred in the term of  $\Delta G_{\text{vdw}}$ , indicating that the difference may be mainly attributed in the aglycone side chain instead of the competitive binding site. It indicates that although NSC679207 has higher binding free energy than that of canagliflozin, it is still sufficient to serve as a novel *N*-glycoside hSGLT2 inhibitor. On the other hand, TCM\_Piperenol\_A (non-glycoside compound) showed great potential to be a promising non-glycoside hSGLT2 inhibitor

**Table 7.** Binding free energies and their decomposed energy components for canagliflozin, NSC679207 and TCM\_Piperenol\_A using the MM/PBSA method.

Compound	$\Delta G_{\text{vdw}}$	$\Delta G_{\text{elec}}$	$\Delta G_{\text{polar}}$	$\Delta G_{\text{nonpolar}}$	$\Delta G_{\text{binding}}$
Canagliflozin	−47.01	−20.90	54.40	−38.81	−52.32
NSC679207	−42.63	−20.05	51.98	−34.60	−45.30
TCM_Piperenol_A	−50.55	−4.95	36.31	−36.00	−55.19



**Figure 8.** 3D structures and 2D interaction diagrams of canagliflozin, NSC679207 and TCM\_Piperenol\_A in the hSGLT2 binding site. (A) 3D structure of canagliflozin in the hSGLT2 binding site. (B) 2D interaction of canagliflozin with hSGLT2. (C) 3D structure of NSC679207 in the hSGLT2 binding site. (D) 2D interaction of NSC679207 with hSGLT2. (E) 3D structure of TCM\_Piperenol\_A in the hSGLT2 binding site. (F) 2D interaction of TCM\_Piperenol\_A with hSGLT2. Hydrogen bonds in 3D structures are shown in green.

due to its good ADMET properties, binding stability, and better binding free energy than canagliflozin.

To the best of our knowledge, this is the first attempt to employ a combination of a series of computational approaches to discover novel *N*-glycoside and non-glycoside hSGLT2 inhibitors. In this study, a ligand-based pharmacophore containing 3 HBs and 1 HY features was constructed and validated. The subsequent docking results showed that the most important binding affinities of the selected hSGLT2 inhibitors were provided by hydrogen bonding network, which is consistent with the previous studies. TCM is by far the largest free downloadable database for Chinese herbal medicine in the world. In previous studies, the possibility of natural compounds to serve as hSGLT2 inhibitors have been revealed ever since the first SGLTs inhibitor, phlorizin, a substance found in some fruit trees, was discovered [66]. TCM\_Pipereno\_A is isolated from Piper cubeba (cubeb pepper), which is mainly grown in the islands of Java and Sumatra and usually used as a spicy ingredient. Our study establishes a fast and low cost platform to successfully discover novel and *N*-glycoside and non-glycoside hSGLT2 inhibitors from various databases, which may become potential leads for the treatment of T2DM after a series of *in vitro* and/or *in vivo* bioassays in the future.

#### 4. Conclusion

The priority of discovering novel hSGLT2 inhibitors with less side effects has been emphasized and increased year by year for the development of T2DM drugs [67]. However, the development of hSGLT2 inhibitors is still limited to the *C*-glycoside inhibitors in both commercially available drugs and clinical trials. Although the advantages of *N*-glycoside and non-glycoside inhibitors have been reported, the discovery of novel drugs is still high cost and time-consuming. In this study, we used highly accurate pharmacophore model and virtual screening to retrieve various types of compounds from large databases rapidly. In addition to the well established NCI database, the TCM database, which contains abundant Chinese herbal medicines and natural products, was also used as the screened database in this study. Finally, NSC679207 and TCM\_Pipereno\_A, which exhibit different chemical scaffolds than that of canagliflozin, were successfully identified to possess higher binding affinities towards hSGLT2. Besides, they also exhibit good pharmacokinetic properties and low risk in several common lethal side effects via ADMET prediction. Thus, these two compounds exhibit great potential to serve as novel *N*-glycoside and non-glycoside hSGLT2 inhibitors and may be further subjected to a series of *in vitro* and/or *in vivo* bioassays. Finally, our results also open a new channel leading to a prospective development for T2DM therapy.

#### Acknowledgements

The authors greatly thank the Ministry of Science and Technology of Taiwan

(MOST 106-2221-E-027-117-MY3 and MOST 107-2221-E-027-038-MY3) and National Taipei University of Technology and Taipei Medical University joint project (USTP-NTUT-TMU-105-05) for their financial supports.

### Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

### References

- [1] Kharroubi, A.T. and Darwish, H.M. (2015) Diabetes Mellitus: The Epidemic of the Century. *World Journal of Diabetes*, **6**, 850-867. <https://doi.org/10.4239/wjd.v6.i6.850>
- [2] DeFronzo, R.A., Ferrannini, E., Groop, L., Henry, R.R., Herman, W.H. and Holst, J.J. (2015) Type 2 Diabetes Mellitus. *Nature Reviews Disease Primers*, **1**, Article No. 15019. <https://doi.org/10.1038/nrdp.2015.19>
- [3] Rewers, M. (2012) Challenges in Diagnosing Type 1 Diabetes in Different Populations. *Diabetes & Metabolism Journal*, **36**, 90-97. <https://doi.org/10.4093/dmj.2012.36.2.90>
- [4] Dwarakanathan, A. (2006) Diabetes Update. *Journal of Insurance Medicine*, **38**, 20-30.
- [5] Chaudhury, A., Duvoor, C., Reddy Dendi, V.S., Kraleti, S., Chada, A. and Ravilla, R. (2017) Clinical Review of Antidiabetic Drugs: Implications for Type 2 Diabetes Mellitus Management. *Frontiers in Endocrinology (Lausanne)*, **8**, 6. <https://doi.org/10.3389/fendo.2017.00006>
- [6] Inzucchi, S.E. (2002) Oral Antihyperglycemic Therapy for Type 2 Diabetes: Scientific Review. *JAMA*, **287**, 360-372. <https://doi.org/10.1001/jama.287.3.360>
- [7] Marin-Penalver, J.J., Martin-Timon, I., Sevillano-Collantes, C. and Del Canizo-Gomez, F.J. (2016) Update on the Treatment of Type 2 Diabetes Mellitus. *World Journal of Diabetes*, **7**, 354-395. <https://doi.org/10.4239/wjd.v7.i17.354>
- [8] Bakris, G.L., Fonseca, V.A., Sharma, K. and Wright, E.M. (2009) Renal Sodium-Glucose Transport: Role in Diabetes Mellitus and Potential Clinical Implications. *Kidney International*, **75**, 1272-1277. <https://doi.org/10.1038/ki.2009.87>
- [9] Vivian, E.M. (2014) Sodium-Glucose Co-Transporter 2 (SGLT2) Inhibitors: A Growing Class of Antidiabetic Agents. *Drugs Context*, **3**, 212-264. <https://doi.org/10.7573/dic.212264>
- [10] Kanai, Y., Lee, W.S., You, G., Brown, D. and Hediger, M.A. (1994) The Human Kidney Low Affinity Na<sup>+</sup>/Glucose Cotransporter SGLT2. Delineation of the Major Renal Reabsorptive Mechanism for D-Glucose. *Journal of Clinical Investigation*, **93**, 397-404. <https://doi.org/10.1172/JCI116972>
- [11] Scheen, A.J. (2016) SGLT2 Inhibitors: Benefit/Risk Balance. *Current Diabetes Reports*, **16**, 92. <https://doi.org/10.1007/s11892-016-0789-4>
- [12] Wang, Z., Sun, J., Han, R., Fan, D., Dong, X. and Luan, Z. (2018) Efficacy and Safety of Sodium-Glucose Cotransporter-2 Inhibitors versus Dipeptidyl Peptidase-4 Inhibitors as Monotherapy or Add-On to Metformin in Patients with Type 2 Diabetes Mellitus: A Systematic Review and Meta-Analysis. *Diabetes, Obesity and Metabolism*, **20**, 113-120. <https://doi.org/10.1111/dom.13047>
- [13] Alicic, R.Z., Johnson, E.J. and Tuttle, K.R. (2018) SGLT2 Inhibition for the Preven-

- tion and Treatment of Diabetic Kidney Disease: A Review. *American Journal of Kidney Diseases*, **72**, 267-277. <https://doi.org/10.1053/j.ajkd.2018.03.022>
- [14] Bajaj, H.S., Brown, R.E., Bhullar, L., Sohi, N., Kalra, S. and Aronson, R. (2018) SGLT2 Inhibitors and Incretin Agents: Associations with Alanine Aminotransferase Activity in Type 2 Diabetes. *Diabetes & Metabolism*, **44**, 493-499. <https://doi.org/10.1016/j.diabet.2018.08.001>
- [15] Uthman, L., Baartscheer, A., Bleijlevens, B., Schumacher, C.A., Fiolet, J.W.T. and Koeman, A. (2018) Class Effects of SGLT2 Inhibitors in Mouse Cardiomyocytes and Hearts: Inhibition of Na(+)/H(+) Exchanger, Lowering of Cytosolic Na(+) and Vasodilation. *Diabetologia*, **61**, 722-726. <https://doi.org/10.1007/s00125-017-4509-7>
- [16] Lin, J.T., Hahn, K.D. and Kinne, R. (1982) Synthesis of Phlorizin Derivatives and Their Inhibitory Effect on the Renal Sodium/D-Glucose Cotransport System. *Biochimica et Biophysica Acta*, **693**, 379-388. [https://doi.org/10.1016/0005-2736\(82\)90445-X](https://doi.org/10.1016/0005-2736(82)90445-X)
- [17] Ehrenkranz, J.R., Lewis, N.G., Kahn, C.R. and Roth, J. (2005) Phlorizin: A Review. *Diabetes/Metabolism Research and Reviews*, **21**, 31-38. <https://doi.org/10.1002/dmrr.532>
- [18] Frampton, J.E. (2018) Empagliflozin: A Review in Type 2 Diabetes. *Drugs*, **78**, 1037-1048. <https://doi.org/10.1007/s40265-018-0937-z>
- [19] Plosker, G.L. (2014) Dapagliflozin: A Review of Its Use in Patients with Type 2 Diabetes. *Drugs*, **74**, 2191-2209. <https://doi.org/10.1007/s40265-014-0324-3>
- [20] Deeks, E.D. and Scheen, A.J. (2017) Canagliflozin: A Review in Type 2 Diabetes. *Drugs*, **77**, 1577-1592. <https://doi.org/10.1007/s40265-017-0801-6>
- [21] Washburn, W.N. (2009) Evolution of Sodium Glucose Co-Transporter 2 Inhibitors as Anti-Diabetic Agents. *Expert Opinion on Therapeutic Patents*, **19**, 1485-1499. <https://doi.org/10.1517/13543770903337828>
- [22] Isaji, M. (2011) SGLT2 Inhibitors: Molecular Design and Potential Differences in Effect. *Kidney International Supplements*, **120**, S14-S19. <https://doi.org/10.1038/ki.2010.511>
- [23] Yao, C.H., Song, J.S., Chen, C.T., Yeh, T.K., Hung, M.S. and Chang, C.C. (2011) Discovery of Novel N-beta-D-xylosylindole Derivatives as Sodium-Dependent Glucose Cotransporter 2 (SGLT2) Inhibitors for the Management of Hyperglycemia in Diabetes. *Journal of Medicinal Chemistry*, **54**, 166-178. <https://doi.org/10.1021/jm101072y>
- [24] Ernst, B. and Magnani, J.L. (2009) From Carbohydrate Leads to Glycomimetic Drugs. *Nature Reviews Drug Discovery*, **8**, 661-677. <https://doi.org/10.1038/nrd2852>
- [25] Wu, J.S., Peng, Y.H., Wu, J.M., Hsieh, C.J., Wu, S.H. and Coumar, M.S. (2010) Discovery of Non-Glycoside Sodium-Dependent Glucose Co-Transporter 2 (SGLT2) Inhibitors by Ligand-Based Virtual Screening. *Journal of Medicinal Chemistry*, **53**, 8770-8774. <https://doi.org/10.1021/jm101080v>
- [26] Guo, C., Hu, M., DeOrazio, R.J., Usyatinsky, A., Fitzpatrick, K. and Zhang, Z.J. (2014) The Design and Synthesis of Novel SGLT2 Inhibitors: C-Glycosides with Benzyltriazolopyridinone and Phenylhydantoin as the Aglycone Moieties. *Bioorganic & Medicinal Chemistry*, **22**, 3414-3422. <https://doi.org/10.1016/j.bmc.2014.04.036>
- [27] Cai, W.Q., Jiang, L.L., Xie, Y.F., Liu, Y.Q., Liu, W. and Zhao, G.L. (2015) Design of SGLT2 Inhibitors for the Treatment of Type 2 Diabetes: A History Driven by Biol-

- ogy to Chemistry. *Medicinal Chemistry*, **11**, 317-328.  
<https://doi.org/10.2174/1573406411666150105105529>
- [28] Aguillon, A.R., Mascarello, A., Segretti, N.D., de Azevedo, H.F.Z., Guimaraes, C.R.W. and Miranda, L.S.M. (2018) Synthetic Strategies toward SGLT2 Inhibitors. *Organic Process Research & Development*, **22**, 467-488.  
<https://doi.org/10.1021/acs.oprd.8b00017>
- [29] Neumiller, J.J. (2014) Empagliflozin: A New Sodium-Glucose Co-Transporter 2 (SGLT2) Inhibitor for the Treatment of Type 2 Diabetes. *Drugs Context*, **3**, Article ID: 212262. <https://doi.org/10.7573/dic.212262>
- [30] Fujimori, Y., Katsuno, K., Nakashima, I., Ishikawa-Takemura, Y., Fujikura, H. and Isaji, M. (2008) Remogliflozin Etabonate, in a Novel Category of Selective Low-Affinity Sodium Glucose Cotransporter (SGLT2) Inhibitors, Exhibits Antidiabetic Efficacy in Rodent Models. *Journal of Pharmacology and Experimental Therapeutics*, **327**, 268-276. <https://doi.org/10.1124/jpet.108.140210>
- [31] Miao, Z., Nucci, G., Amin, N., Sharma, R., Mascitti, V. and Tugnait, M. (2013) Pharmacokinetics, Metabolism, and Excretion of the Antidiabetic Agent Ertugliflozin (PF-04971729) in Healthy Male Subjects. *Drug Metabolism & Disposition*, **41**, 445-456. <https://doi.org/10.1124/dmd.112.049551>
- [32] Ohtake, Y., Sato, T., Kobayashi, T., Nishimoto, M., Taka, N. and Takano, K. (2012) Discovery of Tofogliflozin, a Novel C-Arylglucoside with an O-Spiroketal Ring System, as a Highly Selective Sodium Glucose Cotransporter 2 (SGLT2) Inhibitor for the Treatment of Type 2 Diabetes. *Journal of Medicinal Chemistry*, **55**, 7828-7840.  
<https://doi.org/10.1021/jm300884k>
- [33] Xu, B., Feng, Y., Lv, B., Xu, G., Zhang, L. and Du, J. (2010) Ortho-Substituted C-Aryl Glucosides as Highly Potent and Selective Renal Sodium-Dependent Glucose Co-Transporter 2 (SGLT2) Inhibitors. *Bioorganic & Medicinal Chemistry*, **18**, 4422-4432. <https://doi.org/10.1016/j.bmc.2010.04.088>
- [34] Yamamoto, Y., Kawanishi, E., Koga, Y., Sakamaki, S., Sakamoto, T. and Ueta, K. (2013) N-Glucosides as Human Sodium-Dependent Glucose Cotransporter 2 (hSGLT2) Inhibitors. *Bioorganic & Medicinal Chemistry Letters*, **23**, 5641-5645.  
<https://doi.org/10.1016/j.bmcl.2013.08.042>
- [35] Zhang, X., Urbanski, M., Patel, M., Zeck, R.E., Cox, G.G. and Bian, H. (2005) Heteroaryl-O-Glucosides as Novel Sodium Glucose Co-Transporter 2 Inhibitors. Part 1. *Bioorganic & Medicinal Chemistry Letters*, **15**, 5202-5206.  
<https://doi.org/10.1016/j.bmcl.2005.08.067>
- [36] Zhang, W., Welihinda, A., Mechanic, J., Ding, H., Zhu, L. and Lu, Y. (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. *Pharmacological Research*, **63**, 284-293. <https://doi.org/10.1016/j.phrs.2011.01.001>
- [37] Tang, C., Zhu, X., Huang, D., Zan, X., Yang, B. and Li, Y. (2012) A Specific Pharmacophore Model of Sodium-Dependent Glucose Co-Transporter 2 (SGLT2) Inhibitors. *Journal of Molecular Modeling*, **18**, 2795-2804.  
<https://doi.org/10.1007/s00894-011-1303-1>
- [38] John, S., Thangapandian, S., Sakkiah, S. and Lee, K.W. (2011) Potent BACE-1 Inhibitor Design Using Pharmacophore Modeling, in Silico Screening and Molecular Docking Studies. *BMC Bioinformatics*, **12**, S28.  
<https://doi.org/10.1186/1471-2105-12-S1-S28>
- [39] Lipinski, C.A., Lombardo, F., Dominy, B.W. and Feeney, P.J. (2001) Experimental and Computational Approaches to Estimate Solubility and Permeability in Drug

- Discovery and Development Settings. *Advanced Drug Delivery Reviews*, **46**, 3-26. [https://doi.org/10.1016/S0169-409X\(00\)00129-0](https://doi.org/10.1016/S0169-409X(00)00129-0)
- [40] Veber, D.F., Johnson, S.R., Cheng, H.Y., Smith, B.R., Ward, K.W. and Kopple, K.D. (2002) Molecular Properties That Influence the Oral Bioavailability of Drug Candidates. *Journal of Medicinal Chemistry*, **45**, 2615-2623. <https://doi.org/10.1021/jm020017n>
- [41] Watanabe, A., Choe, S., Chaptal, V., Rosenberg, J.M., Wright, E.M. and Grabe, M. (2010) The Mechanism of Sodium and Substrate Release from the Binding Pocket of vSGLT. *Nature*, **468**, 988-991. <https://doi.org/10.1038/nature09580>
- [42] Ho, B.K. and Brasseur, R. (2005) The Ramachandran Plots of Glycine and Pre-Proline. *BMC Structural Biology*, **5**, 14. <https://doi.org/10.1186/1472-6807-5-14>
- [43] Bodade, R.G., Beedkar, S.D., Manwar, A.V. and Khobragade, C.N. (2010) Homology Modeling and Docking Study of Xanthine Oxidase of *Arthrobacter* sp. XL26. *International Journal of Biological Macromolecules*, **47**, 298-303. <https://doi.org/10.1016/j.ijbiomac.2010.04.002>
- [44] Faham, S., Watanabe, A., Besserer, G.M., Cascio, D., Specht, A. and Hirayama, B.A. (2008) The Crystal Structure of a Sodium Galactose Transporter Reveals Mechanistic Insights into Na<sup>+</sup>/Sugar Symport. *Science*, **321**, 810-814. <https://doi.org/10.1126/science.1160406>
- [45] Gao, J., Zhang, Q., Liu, M., Zhu, L., Wu, D. and Cao, Z. (2016) bSiteFinder, an Improved Protein-Binding Sites Prediction Server Based on Structural Alignment: More Accurate and Less Time-Consuming. *Journal of Cheminformatics*, **8**, 38. <https://doi.org/10.1186/s13321-016-0149-z>
- [46] Wu, G., Robertson, D.H., Brooks, C.L. and Vieth, M. (2003) Detailed Analysis of Grid-Based Molecular Docking: A Case Study of CDOCKER-A CHARMM-Based MD Docking Algorithm. *Journal of Computational Chemistry*, **24**, 1549-1562. <https://doi.org/10.1002/jcc.10306>
- [47] Cheng, F., Li, W., Zhou, Y., Shen, J., Wu, Z. and Liu, G. (2012) admetSAR: A Comprehensive Source and Free Tool for Assessment of Chemical ADMET Properties. *Journal of Chemical Information and Modeling*, **52**, 3099-3105. <https://doi.org/10.1021/ci300367a>
- [48] Van Der Spoel, D., Lindahl, E., Hess, B., Groenhof, G., Mark, A.E. and Berendsen, H.J. (2005) GROMACS: Fast, Flexible, and Free. *Journal of Computational Chemistry*, **26**, 1701-1718. <https://doi.org/10.1002/jcc.20291>
- [49] Lindorff-Larsen, K., Piana, S., Palmo, K., Maragakis, P., Klepeis, J.L. and Dror, R.O. (2010) Improved Side-Chain Torsion Potentials for the Amber ff99SB Protein Force Field. *Proteins*, **78**, 1950-1958. <https://doi.org/10.1002/prot.22711>
- [50] Wang, J., Wolf, R.M., Caldwell, J.W., Kollman, P.A. and Case, D.A. (2004) Development and Testing of a General Amber Force Field. *Journal of Computational Chemistry*, **25**, 1157-1174. <https://doi.org/10.1002/jcc.20035>
- [51] Kumari, R., Kumar, R., Open Source Drug Discovery C and Lynn, A. (2014) G\_mmpbsa—A GROMACS Tool for High-Throughput MM-PBSA Calculations. *Journal of Chemical Information and Modeling*, **54**, 1951-1962. <https://doi.org/10.1021/ci500020m>
- [52] Kandakatla, N. and Ramakrishnan, G. (2014) Ligand Based Pharmacophore Modeling and Virtual Screening Studies to Design Novel HDAC2 Inhibitors. *Advances in Bioinformatics*, **2014**, Article ID: 812148. <https://doi.org/10.1155/2014/812148>
- [53] Agarwal, A., Paliwal, S., Mishra, R., Sharma, S., Kumar Dwivedi, A. and Tripathi, R.

- (2015) Discovery of a Selective, Safe and Novel Anti-Malarial Compound with Activity against Chloroquine Resistant Strain of *Plasmodium falciparum*. *Scientific Reports*, **5**, Article No. 13838. <https://doi.org/10.1038/srep13838>
- [54] Yu, S.L., Yuan, J.T., Zhang, Y., Gao, S.F., Gan, Y. and Han, M. (2017) Combined HQSAR, Topomer CoMFA, Homology Modeling and Docking Studies on Triazole Derivatives as SGLT2 Inhibitors. *Future Medicinal Chemistry*, **9**, 847-858. <https://doi.org/10.4155/fmc-2017-0002>
- [55] Nakka, S. and Guruprasad, L. (2012) Structural Insights into the Active Site of Human Sodium Dependent Glucose Co-Transporter 2: Homology Modelling, Molecular Docking, and 3D-QSAR Studies. *Australian Journal of Chemistry*, **65**, 1314-1324. <https://doi.org/10.1071/CH12051>
- [56] Dong, L., Feng, R., Bi, J., Shen, S., Lu, H. and Zhang, J. (2018) Insight into the Interaction Mechanism of Human SGLT2 with Its Inhibitors: 3D-QSAR Studies, Homology Modeling, and Molecular Docking and Molecular Dynamics Simulations. *Journal of Molecular Modeling*, **24**, 86. <https://doi.org/10.1007/s00894-018-3582-2>
- [57] Liu, W., Wang, H.J. and Meng, F.C. (2015) In Silico Modeling of Aspalathin and Nothofagin against SGLT2. *Journal of Theoretical & Computational Chemistry*, **14**, Article ID: 1550056. <https://doi.org/10.1142/S021963361550056X>
- [58] Xu, J.X., Yuan, H.L., Ran, T., Zhang, Y.M., Liu, H.C. and Lu, S. (2015) A Selectivity Study of Sodium-Dependent Glucose Cotransporter 2/Sodium-Dependent Glucose Cotransporter 1 Inhibitors by Molecular Modeling. *Journal of Molecular Recognition*, **28**, 467-479. <https://doi.org/10.1002/jmr.2464>
- [59] Clive, D. (1985) Mutagenicity in Drug Development: Interpretation and Significance of Test Results. *Regulatory Toxicology and Pharmacology*, **5**, 79-100. [https://doi.org/10.1016/0273-2300\(85\)90022-4](https://doi.org/10.1016/0273-2300(85)90022-4)
- [60] Kalyaanamoorthy, S. and Barakat, K.H. (2018) Binding Modes of hERG Blockers: An Unsolved Mystery in the Drug Design Arena. *Expert Opinion on Drug Discovery*, **13**, 207-210. <https://doi.org/10.1080/17460441.2018.1418319>
- [61] Rizzo, N., Hispard, E., Dolbeault, S., Dally, S., Leverge, R. and Girre, C. (1997) Impact of Long-Term Ethanol Consumption on CYP1A2 Activity. *Clinical Pharmacology & Therapeutics*, **62**, 505-509. [https://doi.org/10.1016/S0009-9236\(97\)90045-X](https://doi.org/10.1016/S0009-9236(97)90045-X)
- [62] Holmes, M.V., Perel, P., Shah, T., Hingorani, A.D. and Casas, J.P. (2011) CYP2C19 Genotype, Clopidogrel Metabolism, Platelet Function, and Cardiovascular Events: A Systematic Review and Meta-Analysis. *JAMA*, **306**, 2704-2714. <https://doi.org/10.1001/jama.2011.1880>
- [63] Cheng, F.X., Yu, Y., Zhou, Y.D., Shen, Z.H., Xiao, W. and Liu, G.X. (2011) Insights into Molecular Basis of Cytochrome P450 Inhibitory Promiscuity of Compounds. *Journal of Chemical Information and Modeling*, **51**, 2482-2495. <https://doi.org/10.1021/ci200317s>
- [64] Hornak, V., Abel, R., Okur, A., Strockbine, B., Roitberg, A. and Simmerling, C. (2006) Comparison of Multiple Amber Force Fields and Development of Improved Protein Backbone Parameters. *Proteins*, **65**, 712-725. <https://doi.org/10.1002/prot.21123>
- [65] Ghezzi, C., Loo, D.D.F. and Wright, E.M. (2018) Physiology of Renal Glucose Handling via SGLT1, SGLT2 and GLUT2. *Diabetologia*, **61**, 2087-2097. <https://doi.org/10.1007/s00125-018-4656-5>
- [66] Choi, C.I. (2016) Sodium-Glucose Cotransporter 2 (SGLT2) Inhibitors from Natural Products: Discovery of Next-Generation Antihyperglycemic Agents. *Molecules*,

**21**, 1136. <https://doi.org/10.3390/molecules21091136>

- [67] Mosley, J.F., Smith, L., Everton, E. and Fellner, C. (2015) Sodium-Glucose Linked Transporter 2 (SGLT2) Inhibitors in the Management of Type-2 Diabetes: A Drug Class Overview. *PT*, **40**, 451-462.