Supporting Information

DIA-DB: A Database and Web Server for the Prediction of Diabetes Drugs

Horacio Pérez-Sánchez[1,*], Helena den-Haan[1,2], Jorge Peña-García[1], Jesús Lozano-Sánchez[3,4], María Encarnación Martínez Moreno[1], Antonia Sánchez-Pérez[1], Andrés Muñoz[1], Pedro Ruiz-Espinosa[5], Andreia S.P. Pereira[6], Antigoni Katsikoudi[7], José Antonio Gabaldón Hernández[1], Ivana Stojanovic[8], Antonio Segura Carretero[3,4,*] and Andreas G. Tzakos[7,*].

[1] *Bioinformatics and High Performance Computing Research Group (BIO-HPC), Computer Engineering Department, Universidad Católica de Murcia (UCAM), Guadalupe, 30107, Spain.*

[2] *Villapharma Research S.L., Parque Tecnológico de Fuente Álamo. Ctra. El Estrecho-Lobosillo, Km. 2.5, Av. Azul 30320 Fuente Álamo de Murcia, Murcia, Spain.*

[3] *Research and Development of Functional Food Centre (CIDAF), PTS Granada, Avda. Del Conocimiento s/n, Edificio BioRegión, 18016 Granada, Spain.*

[4] *Department of Analytical Chemistry, University of Granada. Avda. Fuentenueva s/n, 18071, Granada, Spain.*

[5] *Herbafor SL, Calle Purísima, 52, 30620 Fortuna, Murcia, Spain.*

[6] *Department of Biochemistry, Genetics and Microbiology, University of Pretoria, Pretoria Hillcrest 0083, South Africa.*

[7] *Department of Chemistry, University of Ioannina, Ioannina, 45110, Greece.*

[8] *Department of Immunology, Institute for Biological Research "Sinisa Stankovic", University of Belgrade, Bulevar despota Stefana 142,11060 Belgrade, Serbia.*

[*] Address correspondence to these authors at Universidad Católica de Murcia, Spain; Tel: +34-968278819; Email: hperez@ucam.edu; University of Granada, Spain; Tel: +34-958248435; Email: ansegura@ugr.es; University of Ioannina, Greece; Tel: +302651007194; Email: atzakos@uoi.gr.

MATERIALS AND METHODS

1. The Drugs table. This table is composed of all the FDA approved anti-diabetic drugs, which are widely used for the treatment of T2D as well as the most promising experimental anti-diabetic compounds to date (Table S1). For the assembly of the approved anti-diabetic medication, the WHO Collaborating Centre for Drug Statistics Methodology website was used[1]. All the anti-diabetic drugs share the same ATC code - A10 (drugs used in diabetes). There are 3 different categories of A10 drugs: insulins and analogues (A10A), blood glucose lowering drugs, excluding insulins (A10B) and other drugs used in diabetes (A10X). Our database is based mostly on A10B drugs (the A10X category includes only the drug tolrestat). Extended information for each drug was extracted and added to DIA-DB from the DrugBank database[2]. In contrast, the PDB database[3] was utilized for searching of experimental anti-diabetic agents, whereas their properties were obtained from ChemSpider[4] and PubChem[5] databases.

For each compound we have assembled the most important information concerning their bioactivity. The given information about the desired compounds includes: the trade name; the chemical name; the chemical structure; the SMILES code; notes (depends on the drug); the type of diabetes it targets (T1D or T2D); the ligand ID in DrugBank, PubChem and ZINC databases and the protein target.

The user can access the Drugs table and query properties of each compound on the SEARCH page (http://bio-hpc.ucam.edu/dia-db/web/Search/Search.php), by inserting the name, the SMILES code of the desired compound or even the ZINC, DrugBank, PubChem or CHEMBL IDs (Figure S1).

2. The Calculations table. The DIA-DB features two different types of calculations: structure similarity comparison and docking experiments. With the structure similarity comparison, the desired compound will be compared with other known anti-diabetic compounds and the global shape similarity between them will be measured (ligand-based virtual screening, LBVS). In the docking experiments the interaction between the desired compound and a series of anti-diabetic protein targets will be calculated (structure-based virtual screening, SBVS).

The LBVS approach is based on two methods. A global three-dimensional shape similarity search is performed by means of WEGA software[6] by pure shape scoring function comparison, where all parameters are set as default. Additionally, a hybrid similarity search method based on 3D global shape and pharmacophore fitting can be performed by using SHAFTS tool[7]. All ligands in the database were prepared for LBVS using molconvert tool (ChemAxon, Budapest, Hungary), for conformer generation. Molecular parameters for the ligands in the database were calculated by removing salts and neutralizing their protonation state, computing partial charges by MMFF94 force field, adding hydrogen atoms and minimizing energies (default parameters)[8] by using AmberTools (AMBER 2017; University of California, San Francisco)[9]. A maximum of 100 conformations for every compound stored in DIA-DB were generated by means of Omega[10] keeping default parameters. For 12 compounds of the database, stereoisomerism and strict atom typing were set to false so conformations could be generated. Conformation generation for compounds submitted by DIA-DB users are generated by means of Cyndi[11], using default parameters and MMFF94 force field.

The SBVS approach involves the *in silico* docking of the input compound to selected pharmaceutical protein targets implicated in diabetes (Table S2). The crystallized structures of protein targets (PDB codes in Table S2) were obtained from the PDB[3]. Docking calculations are

carried out using Autodock Vina, setting a grid of 25 Å3 centred in the binding site as determined by the position of the crystallized ligand of every protein target and default parameters. Protein targets and ligands are set for docking by means of AutoDock tools as in Forli *et al*., 2006[12].

3. Experimental workflow. Running a Calculations-based prediction using the DIA-DB is a very straightforward procedure (Figure S2). On the Submit Calculations page, the user has to submit either the SMILES code of the desired compound, or the structure of the molecule by using the DRAW MOLECULE option. If so, another window opens, which can be used for the design of the molecule. Next, the user inserts his E-mail and presses the SUBMIT DATA tab. After the chosen experiment is finished, the user will receive an E-mail containing an access code for that specific experiment. By inserting his E-mail address and the received code in the RESULTS page, the user will be able to access the results.

Once the similarity calculations are completed, the scoring function yields a score that ranks from [0-1], ranging from low to 1 pure shape similarity. This score is used for ranking the compounds according to their predicted similarity from higher to lower score. Scoring values are presented in a table, along with 3D modelling showing the alignment of both the query compound and its match, in DIA-DB using Jmol[13].

After the completion of the docking calculations, the user can access a report, automatically generated by the web server, where the compounds are ranked against all the different protein targets in terms of scoring function value. Using PoseView[14], a 2D diagram is offered of the top docking pose of the query compound in the binding site of the pharmaceutical protein target, pinpointing the amino acids that are implicated in direct interaction with the query compound. Additionally, users can access a 3D view of the predicted complex generated by means of PLIP

tool[15]. Through this information, a user can link the relevant finding to literature data on the importance of the interaction determined.

4. Chemicals and apparatus. All chemicals used in this work for the characterization of the *Sclerocarya birrea* stem-bark aqueous extract were of analytical reagent grade. Methanol was purchased from Panreac (Barcelona, Spain), and acetic acid from Fluka and Sigma-Aldrich (Steinheim, Germany). Double-deionized water, with conductivity lower than 18.2MΩ, was obtained with a Milli-Q system (Millipore, Bedford, MA, USA).

5. Sample extraction. *Sclerocarya birrea* stem-barks were provided by Herbafor SL (Murcia, Spain). These samples were collected from Dakar, Sénégal, in 2016. Stem-barks were removed from different parts of the tree trunk manually and transported in darkness. They were air-dried and then grounded with a Retsch ZM200 ultra centrifugal mill that conferred a final powder with a fineness < 40 μm. The obtained *Sclerocarya birrea* stem-bark powder was kept in darkness until needed. Isolation of bioactive compounds from *Sclerocarya birrea* stem-bark powder was performed by shaking the sample powder (6 g) for 90 min with 30 mL of water. The samples were subsequently centrifuged at 13000g for 10 min to remove the solid fraction, and supernatants were collected. Finally, the extract was filtered through a 0.2 μm syringe filter and the solvent was evaporated under vacuum, and kept at -20°C prior to analysis. This procedure was carried out in triplicate.

6. Analytical characterization. The *Sclerocarya birrea* stem-bark aqueous extract was characterized by high-performance liquid chromatography coupled to electro-spray time-of-flight mass spectrometry (HPLC-ESI-TOF/MS). The extract was dissolved in water at concentration of 10000 μg/mL. The HPLC-ESI-TOF/MS method was performed in an Agilent 1200−HPLC system (Agilent Technologies, Waldbronn, Germany), of the Series Rapid Resolution equipped with a

vacuum degasser, autosampler, a binary pump, and a diode−array detector (DAD). The chromatographic analysis was performed according to Jiménez-Sánchez *et al*., 2015[16] with some modifications. The chromatographic separation was carried out in a Zorbax Eclipse Plus RP−C18 analytical column 150 x 4.6 mm i.d., 1.8 μm particle size (Agilent Technologies, Palo Alto, CA, USA). The mobile phase used was water with 0.5% acetic acid as eluent A, and methanol as eluent B. The flow rate was 0.5 mL/min, and the total run time was 37 min using the following multistep linear gradient: 0 min, 5%B for 5 min, 25%B for 8 min, 27% B for 12 min, 30% B for 15 min, 33% B for 17 min, 35% B for 27 min, 55%B for 30 min, 95% B for 33 min, 5%B and finally a conditioning cycle of 5 min with the same conditions for the next analysis. The temperature of the column was maintained at 25°C. The injection volume in the HPLC was 20 μL. The compounds separated were monitored with a mass-spectrometry detector.

The MS analysis was performed using the microTOF (Bruker Daltonik, Bremen, Germany), which was coupled to the HPLC system. At this stage, the use of a splitter was required for the coupling with the MS detector, as the flow arriving to the TOF detector had to be 0.2 mL/min in order to ensure reproducible results and stable spray. The TOF mass spectrometer was equipped with an ESI interface (model G1607A from Agilent Technologies, Palo Alto, CA, USA) operating in negative ion mode. External mass-spectrometer calibration was performed with sodium acetate clusters (5 mM sodium hydroxide in water/2-propanol 1:1 (v/v), with 0.2% of acetic acid) in quadratic high-precision calibration (HPC) regression mode. The calibration solution was injected at the beginning of the run, and all the spectra were calibrated prior to the identification. The optimum values of source and transfer parameters were established according to Lozano-Sánchez *et al*., 2010[17].

The MS data were processed through Data Analysis 4.0 software (Bruker Daltonik), that provided a list of possible elemental formulas using the Generate Molecular Formula Editor. The latter uses a CHNO algorithm providing standard functionalities such as minimum/maximum elemental range, electron configuration, and ring-plus double-bond equivalent, as well as a sophisticated comparison of the theoretical with the measured isotopic pattern (Sigma-Value) for increased confidence in the suggested molecular formula. For confirmation of elemental compositions, the accuracy threshold was established at 10 ppm, for most of the compounds.

RESULTS

1. Analytical qualitative characterization of phenolic and other polar compounds in *Sclerocarya birrea* stem-bark aqueous extract.

The chromatogram of a representative aqueous extract is shown in Figure S4. The tentatively identified phenolic and other polar compounds are summarized in Table S3, including retention times, *m/z* ratios and molecular formula together with their proposed identities. Phenolic compounds were identified by the interpretation of the information generated by TOF analyzer and previously reported literature[16,18,19].

A total of 33 compounds were characterized in the aqueous extract, 13 were unknown (Uk 1-13). Among these, two were characterized as hydroxybenzoic acids or derivatives, 18 were flavonoids and one, the quinic acid, was not a phenolic molecule. Concerning hydroxybenzoic acids, peaks 2 and 4 had a deprotonated molecule at *m/z* 331 and 169, respectively, and were identified as galloyl glucoside or isomer and gallic acid. Indeed, gallic acid and its derivatives have been previously described as the main simple phenols in *Sclerocarya birrea* stem-bark as well as wood of other trees such as eucalyptus or yellow fir[16,20,21]. With regards to the flavonoids, they were the most

abundant phenolic compounds detected in *Sclerocarya birrea*. Among these compounds, proanthocyanidins were the predominant subclass in the aqueous extracts. A high percentage of proanthocyanidins was characterized by the presence of (epi)catechin or (epi)gallocatechin groups. The pharmaceutical interest in these chemical structures due to their bioactivity on different diseases is also well known and has stimulated multidisciplinary research on the composition of proanthocyanidin in plant sources[20,21].

Peak 7, with experimental *m/z* 305 and molecular formula ($C_{15}H_{14}O7$) generated by TOF analyzer, was tentatively proposed as (epi) gallocatechin or isomer. This compound was previously found in *Sclerocarya birrea* hydro-alcoholic extract[16]. Peak 9 yielded deprotonated molecule at *m/z* 761 was identified as bis(epi)gallocatechin monogallate. Peak 10 had a deprotonated molecule at *m/z* 577. According to the literature, this compound was assigned to a (epi)catechin-(epi)catechin[16]. The spectra generated for peak 12 yielded deprotonated molecule at *m/z* 913, which was attributed to bis(epi)gallocatechin digallate. The spectra generated for peaks 13 and 15 gave the same deprotonated molecule at *m/z* 745 with identical molecular formula ($C_{37}H_{30}O_{17}$). These compounds were proposed as (epi)gallocatechin gallate-(epi)catechin or isomers. Peak 16 and 27 had a deprotonated molecule at *m/z* 289 and 13 and 17.7 min, respectively, being identified as epicatechin or isomers. The presence of these isomers is consistent with the metabolite analysis performed by Jiménez-Sánchez *et al*., 2015[16].

Peak 18 was found to be (epi)gallocatechin gallate-(epi)catechin gallate. Peak 20 had a deprotonated molecule at *m/z* 745 and was identified as (epi)catechin gallate-(epi)catechin. Peaks 21 and 28, with the same experimental *m/z* 457 and molecular formula ($C_{22}H_{18}O_{11}$) generated by TOF analyzer, were tentatively proposed as (epi)gallocatechin gallate or isomers. Peak 24 yielded deprotonated molecule at *m/z* 881 was identified as (epi)catechin gallate-(epi)catechin gallate or

isomer. Peak 26 had a deprotonated molecule at *m/z* 603. This compound was characterized as (epi)catechin glucoside gallate. The most representative flavonoids identified in the aqueous extract were (epi)catechin gallate or isomers (peaks 32 and 33).

Table S2. Number of drugs curated in the DIA-DB Drugs table.

Record type	Number of records
Type II diabetes drugs	69
Type I and II diabetes drugs	11
Experimental T2D compounds	150

Table S3. Pharmaceutical targets of T2D for SBVS in the DIA-DB Experiments database.

Protein target name	PDB Code	Biological function	Desired effect
11-β Hydroxysteroid dehydrogenase type 1 (HSD11B1)	4K1L	Coverts inactive glucocorticoid precursors to active glucocorticoids; glucocorticoids counteract the effects of insulin[22].	Inhibition
Aldose reductase (AKR1B1)	3G5E	Catalyses the reduction of glucose to sorbitol in the polyol pathway, plays a role in diabetic complications[23].	Inhibition
Dipeptidyl peptidase 4 (DPP4)	4A5S	Degrades and inactivates glucagon-like peptide-1 that stimulates insulin secretion from pancreas[24].	Inhibition
Free fatty acid receptor 1 (FFAR1)	4PHU	Binding of free fatty acids to receptor results in increased glucose-stimulated insulin secretion[25].	Activation
Fructose-1,6-bisphosphatase (FBP1)	2JJK	Catalyses the second last step in gluconeogenesis[26].	Inhibition
Glucokinase (GCK)	3IMX	Phosphorylates glucose to glucose-6-phosphate for glycolysis or glycogen synthesis[27].	Activation
Insulin receptor (INSR)	3EKN	Regulates glucose uptake as well as glycogen, lipid and protein synthesis[24].	Activation
Intestinal maltase-glucoamylase (MGAM)	3L4Y	Hydrolyses 1,4-alpha bonds, the last step in the digestion of starch to glucose[28].	Inhibition
Liver glycogen phosphorylase (PYGL)	3DDS	Catalyses the first step of glycogenolysis by the phosphorolysis of glycogen to glucose-1-phosphate[29].	Inhibition
Nuclear receptor liver receptor homolog 1 (NR5A2)	4DOR	Regulates the expression of genes involved in bile acid synthesis, cholesterol synthesis and steroidogenesis[30].	Activation
Pancreatic alpha-amylase (AMY2A)	4GQR	Hydrolyses alpha-1,4-glycosidic bonds of starch during digestion of starch to glucose[28].	Inhibition

Peroxisome proliferator-activated receptor alpha (PPARA)	3FEI	Regulates expression of genes involved in lipid metabolism, in particular, the oxidation of fatty acids as well as lipoprotein assembly and lipid transport[31].	Activation
Peroxisome proliferator-activated receptor delta (PPARD)	3PEQ	Regulates expression of genes involved in fatty acid catabolism[31].	Activation
Peroxisome proliferator-activated receptor gamma (PPARG)	2FVJ	Regulates expression of genes involved in adipogenesis and lipid metabolism particularly fatty acid transport, lipid droplet formation, triacylglycerol metabolism as well as lipolysis of triglycerides[31].	Activation
Pyruvate dehydrogenase kinase 2 (PDK2)	4MPC	Responsible for inactivating the pyruvate dehydrogenase complex that is involved during glucose oxidation[32].	Inhibition
Retinoic acid receptor alpha (RXRA)	1FM9	Heterodimerizes with PPARs thereby initiating gene transcription[31].	Activation
Retinol binding protein 4 (RBP4)	2WR6	Secreted as an adipokine that reduces insulin signalling and promotes gluconeogenesis[33].	Under investigation
Tyrosine-protein phosphatase non-receptor type 9 (PTPN9)	4GE6	Dephosphorylates the insulin receptor thereby reducing insulin sensitivity[27].	Inhibition

Table S4. Proposed phenolic and other polar compounds characterized in *Sclerocarya birrea* stem-bark aqueous extract by HPLC-ESI-TOF-MS.

Peak	RT exp.[min]	Compound Name	Mol.Formula	*m/z* calc. (M-H)	*m/z* meas. (M-H)	Err [ppm]	mSigma
1	3.2	Quinic acid	$C_7H_{12}O_6$	191.0561	191.0563	1.1	3.3
2	4.3	Sucrose	$C_{12}H_{22}O_{11}$	341.1089	341.1084	1.5	7.6
3	5.2	Citric acid	$C_6H_8O_7$	191.0197	191.0196	-0.4	0.6
4	5.9	UK1	$C_{22}H_{22}O_{18}$	573.0733	573.0748	-2.2	19.9
5	6.8	Galloyl glucoside	$C_{13}H_{16}O_{10}$	331.0671	331.0663	2.2	8.3
6	7.8	UK2	$C_{19}H_{26}O_{15}$	493.1199	493.1189	1.9	4.9
7	8.2	Gallic acid	$C_7H_6O_5$	169.0142	169.0145	1.6	3.4
8	8.6	UK3	$C_{18}H_{26}O_{15}$	465.1250	465.1250	-0.2	7.3
9	8.7	UK4	$C_{21}H_{24}O_{12}$	467.1195	467.1189	1.3	3.9
10	9.5	UK5	$C_{12}H_6O_7$	261.0041	261.0076	13.5	26.1
11	9.6	(Epi)gallocatechin or isomer	$C_{15}H_{14}O_7$	305.0667	305.0667	0.1	2.1
12	9.8	UK6	$C_{18}H_{26}O_{13}$	449.1301	449.129	2.4	15.6
13	10.2	Bis(epi)gallocatechin monogallate	$C_{37}H_{30}O_{18}$	761.1359	761.1357	0.3	4.1
14	10.6	(epi)catechin-(epi)catechin	$C_{30}H_{26}O_{12}$	577.1351	577.1326	4.4	7.4
15	10.8	UK7	$C_{18}H_{26}O_{12}$	433.1351	433.1335	3.8	1.9
16	11.5	Bis(epi)gallocatechin digallate	$C_{44}H_{34}O_{22}$	913.1469	913.1428	4.5	15.5
17	11.8	(Epi)gallocatechin gallate (epi)catechin or isomer	$C_{37}H_{30}O_{17}$	745.1410	745.1383	3.6	4.1
18	12.3	UK8	$C_{21}H_{24}O_{11}$	451.1246	451.1228	4	5
19	12.5	(Epi)gallocatechin or isomer	$C_{15}H_{14}O_7$	305.0667	305.0652	4.7	3.2
20	12.6	(Epi)gallocatechin gallate (epi)catechin or isomer	$C_{37}H_{30}O_{17}$	745.1410	745.1382	3.8	4.3
21	13	Catechin	$C_{15}H_{14}O_6$	289.0718	289.0713	1.6	3.2

22	13.2	UK9	$C_{22}H_{34}O_{15}$	537.1825	537.1799	4.9	3.6
23	13.7	(Epi)gallocatechin gallate (epi)catechin gallate	$C_{44}H_{34}O_{21}$	897.1520	897.1484	4	7.1
24	14.0	(Epi)catechin gallate (epi)catechin or isomer	$C_{37}H_{30}O_{16}$	729.1438	729.1440	3.2	13.2
25	14.1	UK10 or isomer	$C_{20}H_{22}O_{12}$	453.1038	453.1008	6.7	5.1
26	14.5	(Epi)catechin gallate (epi)catechin or isomer	$C_{37}H_{30}O_{16}$	729.1461	729.1440	2.9	3.2
27	15.6	(Epi)gallocatechin gallate	$C_{22}H_{18}O_{11}$	457.0776	457.0748	6.2	3.4
28	16.1	UK11	$C_{28}H_{20}O_{7}$	467.1195	467.1162	7	7.1
29	16.2	UK12	$C_{22}H_{26}O_{13}$	497.1301	497.1263	7.5	4.5
30	16.5	(Epi)catechin gallate (epi)catechin gallate or isomer	$C_{44}H_{34}O_{20}$	881.1571	881.1522	5.5	8.5
31	16.9	UK10 or isomer	$C_{20}H_{22}O_{12}$	453.1038	453.1004	7.5	2.8
32	17.2	(Epi)catechin glucoside gallate or isomer	$C_{28}H_{28}O_{15}$	603.1355	603.1314	6.9	4.7
33	17.7	Epicatechin	$C_{15}H_{14}O_{6}$	289.0718	289.0699	6.3	5.1
34	18.5	(Epi)gallocatechin gallate	$C_{22}H_{18}O_{11}$	457.0776	457.0738	8.4	2.5
35	18.7	UK13	$C_{22}H_{26}O_{12}$	481.1351	481.1308	9	2.1
36	19	UK14	$C_{19}H_{26}O_{13}$	461.1301	461.1259	9.1	6.3
37	20	UK15	$C_{22}H_{24}O_{13}$	495.1144	495.1094	10	5
38	21.7	(Epi)catechin gallate or isomer	$C_{22}H_{18}O_{10}$	441.0827	441.0798	6.7	2.5
39	22.7	(Epi)catechin glucoside gallate or isomer	$C_{28}H_{28}O_{15}$	603.1355	603.1308	7.8	12.4
40	23.6	(Epi)catechin gallate or isomer 2	$C_{22}H_{18}O_{10}$	441.0827	441.079	8.5	4.6
41	24.7	UK16	$C_{33}H_{30}O_{7}$	537.1919	537.1917	0.3	38.5
42	25.5	(Epi)afzelechin gallate	$C_{22}H_{18}O_{9}$	425.0878	425.0851	6.4	5.3
43	28.2	UK17	$C_{28}H_{14}O_{8}$	477.0616	477.0642	-5.5	28.3
44	30.3	Ellagic acid	$C_{14}H_{6}O_{8}$	300.9990	300.9985	1.2	3.5

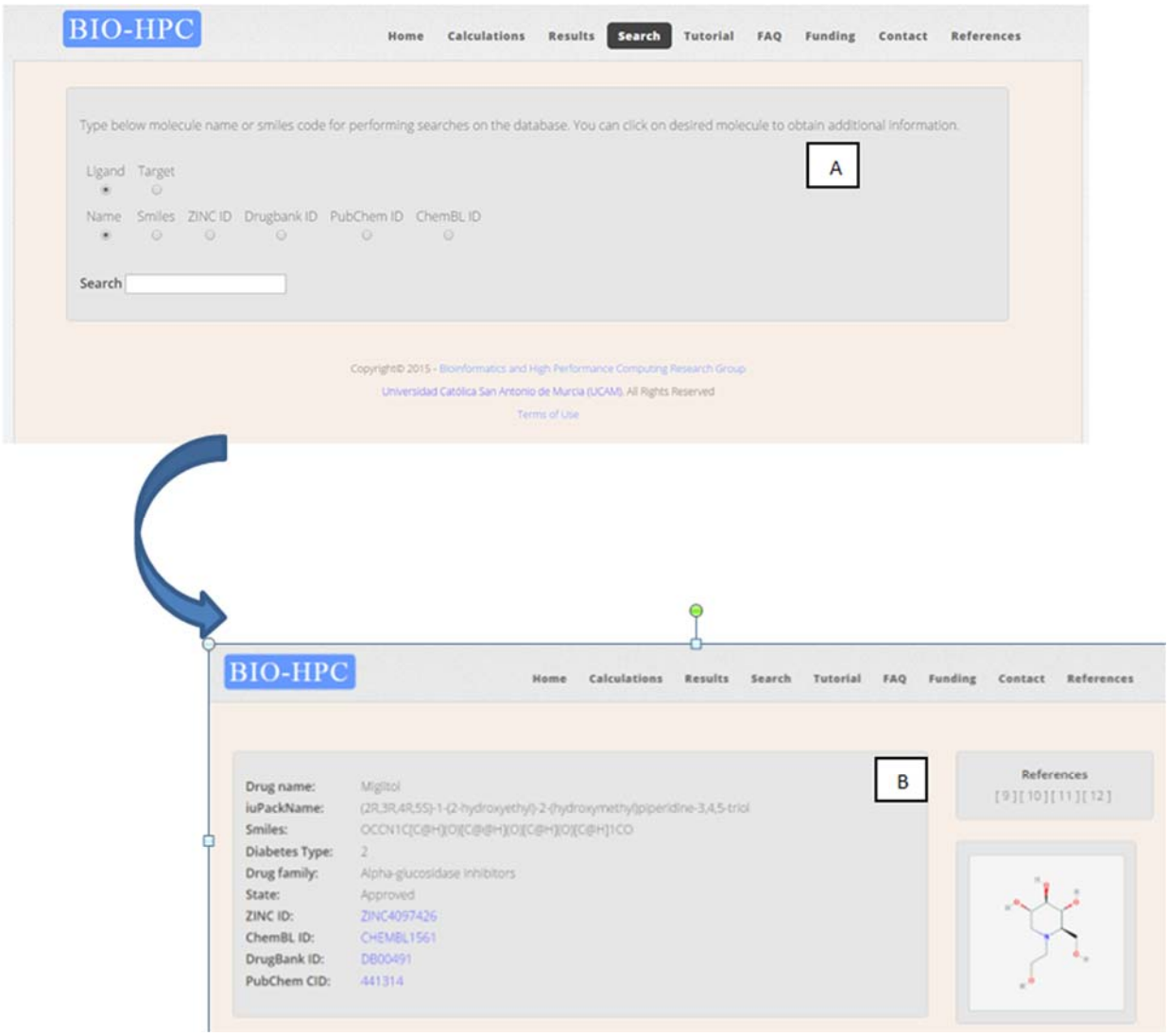


Figure S1. In order to access the database, the user has to select the Search option of the server (A) and type insert the name or SMILES of the desired compound. After that, a table will be presented, containing the sought-after information (B).

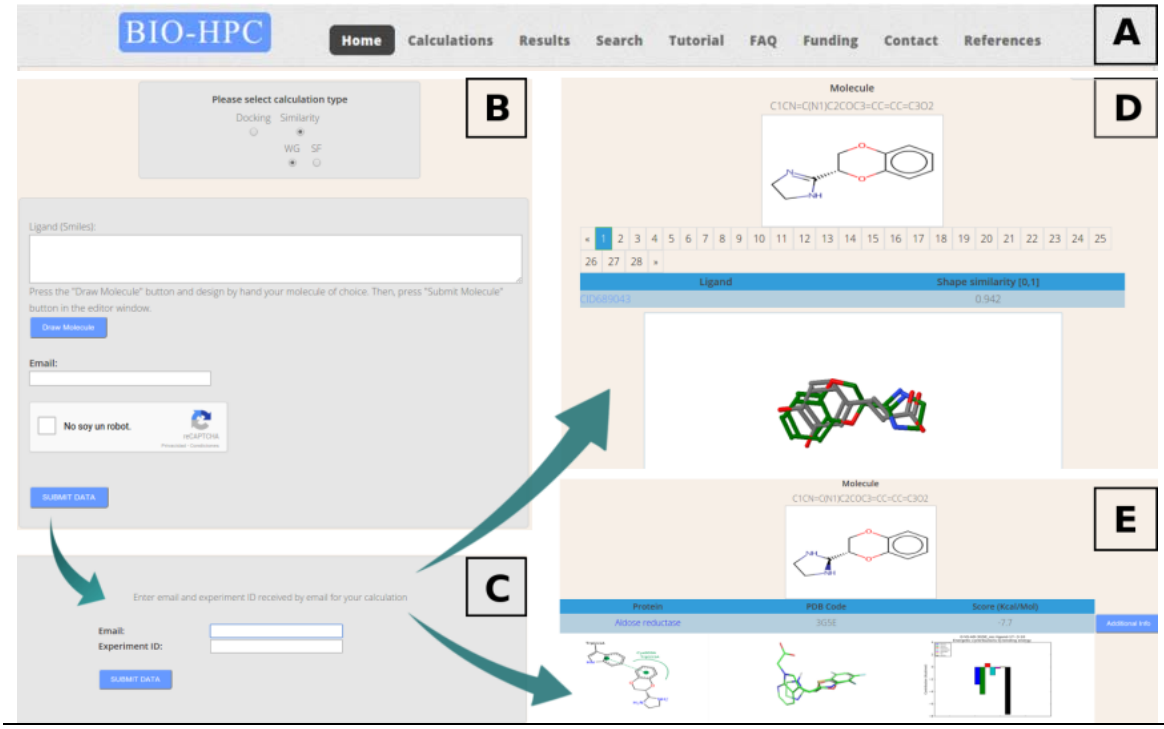


Figure S2. The necessary steps for task submission. At first the user has to access the DIA-DB home page (A) and then choose the Submit Calculations option (B), where either the SMILES or the molecular structure could be inserted. After pressing the Submit Data option and receiving the respective code, the user can access the Results section of the server (C) and insert his e-mail and the required code. By pressing the Submit Data option the results will be shown in a split second (D and E).

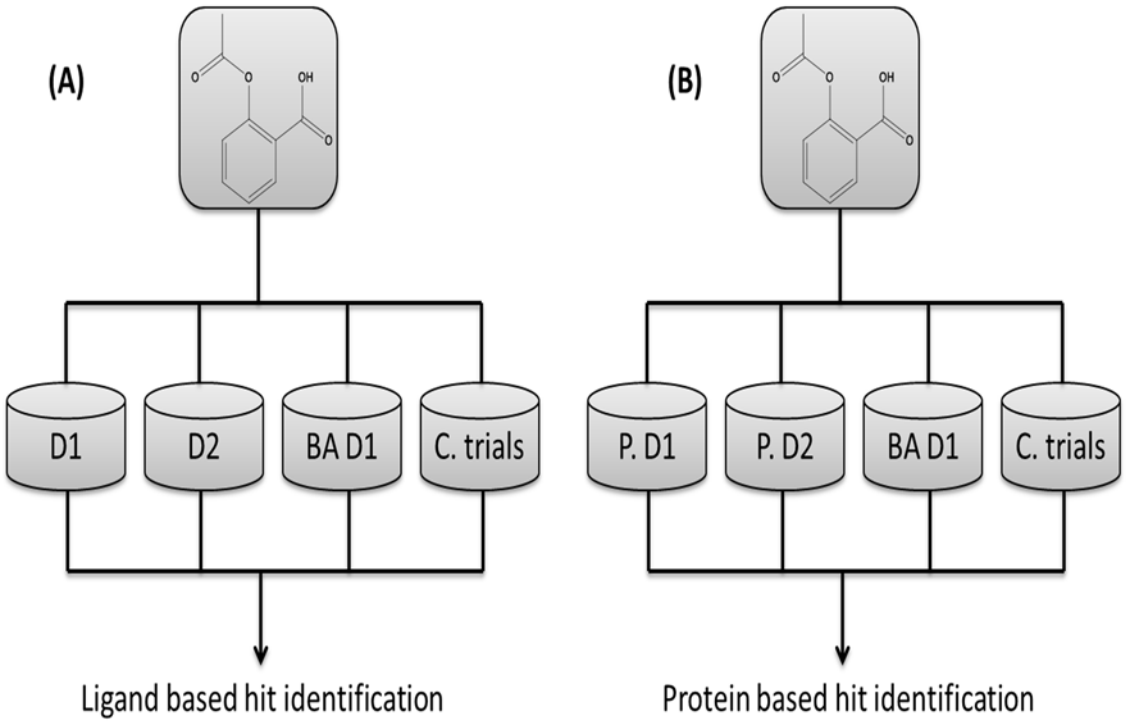


Figure S3. Workflow for diabetes drugs prediction based on: (A) a ligand-based prediction approach and (B) protein-based prediction approach.

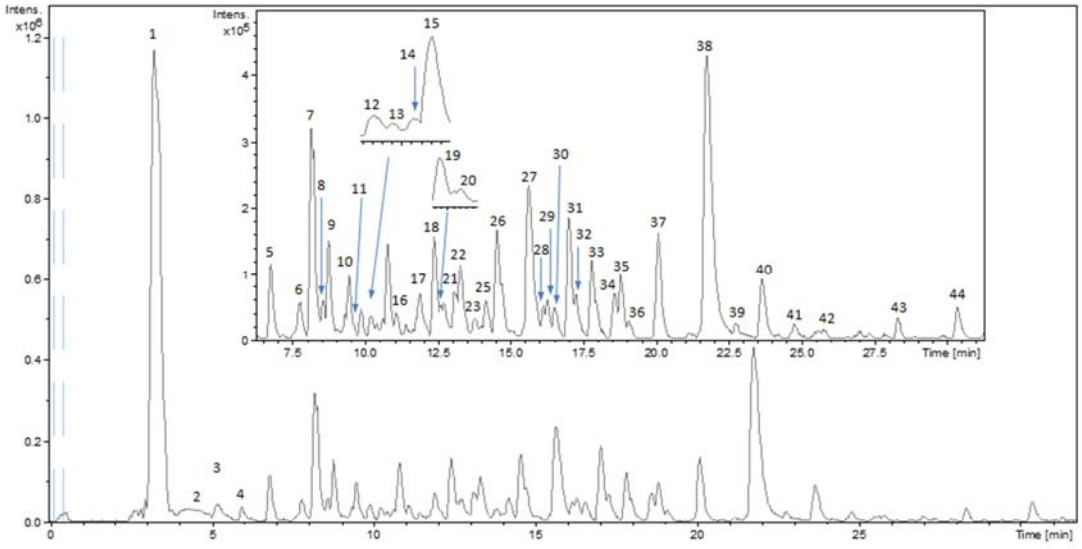
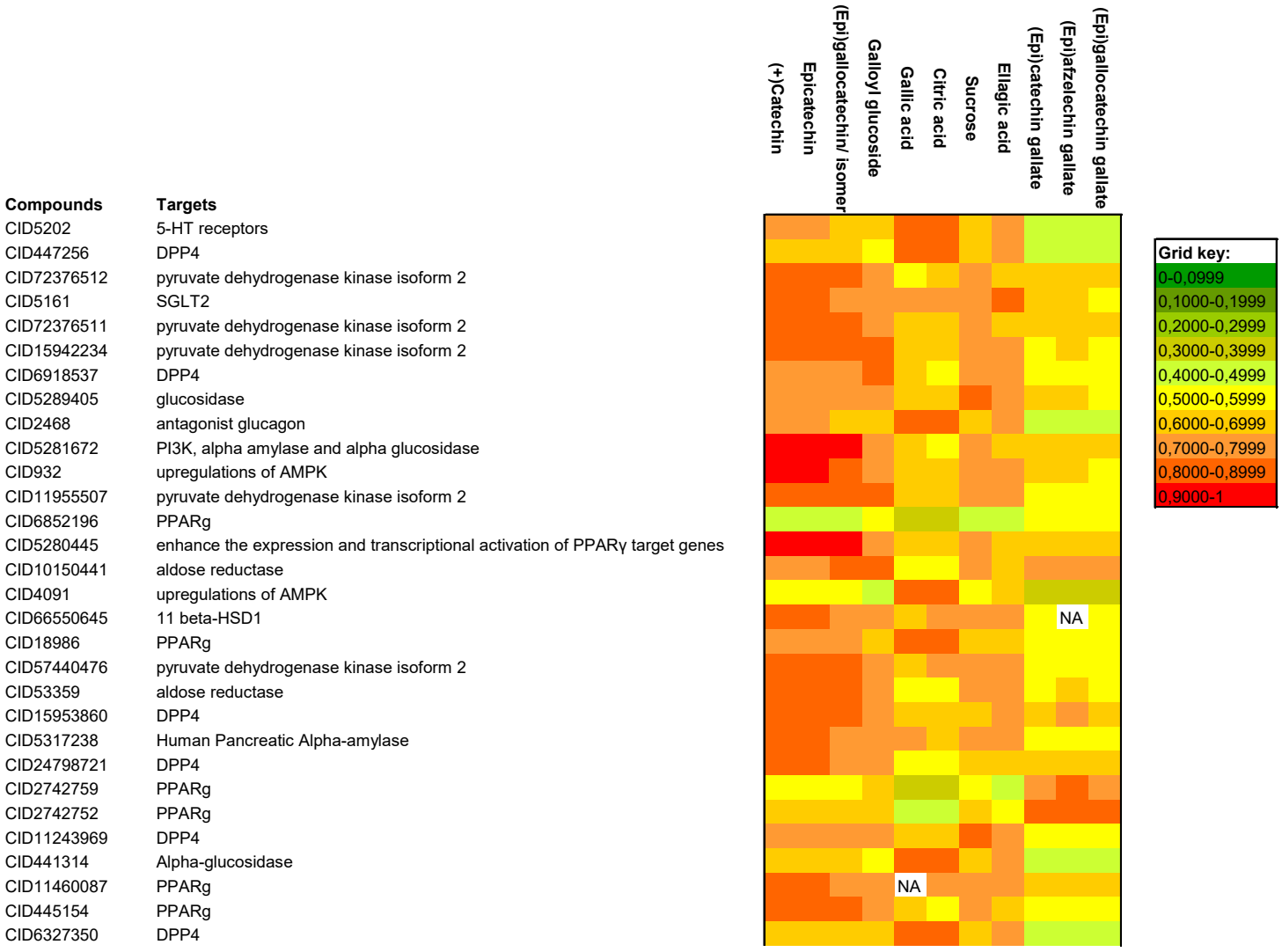


Figure S4. Base Peak Chromatogram of *Sclerocarya birrea* stem-bark aqueous extract. Number of peaks refer to compounds tentatively identified in Table S3.



Compounds	Targets
CID5202	5-HT receptors
CID447256	DPP4
CID72376512	pyruvate dehydrogenase kinase isoform 2
CID5161	SGLT2
CID72376511	pyruvate dehydrogenase kinase isoform 2
CID15942234	pyruvate dehydrogenase kinase isoform 2
CID6918537	DPP4
CID5289405	glucosidase
CID2468	antagonist glucagon
CID5281672	PI3K, alpha amylase and alpha glucosidase
CID932	upregulations of AMPK
CID11955507	pyruvate dehydrogenase kinase isoform 2
CID6852196	PPARg
CID5280445	enhance the expression and transcriptional activation of PPARγ target genes
CID10150441	aldose reductase
CID4091	upregulations of AMPK
CID66550645	11 beta-HSD1
CID18986	PPARg
CID57440476	pyruvate dehydrogenase kinase isoform 2
CID53359	aldose reductase
CID15953860	DPP4
CID5317238	Human Pancreatic Alpha-amylase
CID24798721	DPP4
CID2742759	PPARg
CID2742752	PPARg
CID11243969	DPP4
CID441314	Alpha-glucosidase
CID11460087	PPARg
CID445154	PPARg
CID6327350	DPP4

Figure S5. Similarity results obtained for most relevant compounds of *Sclerocarya birrea* stem-bark aqueous extract.

REFERENCES

(1) World Health Organization Collaborating Centre for Drug Statistics Methodology Website. (http://www.who.int/medicines/regulation/medicines-safety/about/collab-centres-norwegian/en/)

(2) Wishart, D. S.; Knox, C.; Guo, A. C.; Shrivastava, S.; Hassanali, M.; Stothard, P.; Chang, Z.; Woolsey, J., DrugBank: A Comprehensive Resource for In Silico Drug Discovery and Exploration. *Nucleic Acids Res.* **2006**, *34*, D668-D672.

(3) Rose, P. W.; Prlić, A.; Altunkaya, A.; Bi, C.; Bradley, A. R.; Christie, C. H.; Costanzo, L. D.; Duarte, J. M.; Dutta, S.; Feng, Z., The RCSB Protein Data Bank: Integrative View of Protein, Gene and 3D Structural Information. *Nucleic Acids Res.* **2016**, *45*, D271-D281.

(4) Pence, H. E.; Williams, A., ChemSpider: An Online Chemical Information Resource. *J. Chem. Educ.* **2010**, *87*, 1123-1124.

(5) Kim, S.; Thiessen, P. A.; Bolton, E. E.; Chen, J.; Fu, G.; Gindulyte, A.; Han, L.; He, J.; He, S.; Shoemaker, B. A., PubChem Substance and Compound Databases. *Nucleic Acids Res.* **2015**, *44*, D1202-D1213.

(6) Yan, X.; Li, J.; Liu, Z.; Zheng, M.; Ge, H.; Xu, J., Enhancing Molecular Shape Comparison by Weighted Gaussian Functions. *J. Chem. Inf. Model.* **2013**, *53*, 1967-1978.

(7) Liu, X.; Jiang, H.; Li, H., SHAFTS: A Hybrid Approach for 3D Molecular Similarity Calculation. 1. Method and Assessment of Virtual Screening. *J. Chem. Inf. Model.* **2011**, *51*, 2372-2385.

(8) Halgren, T. A., Potential Energy Functions. *Curr. Opin. Struct. Biol.* **1995**, *5*, 205-210.

(9) Case, D.; Cerutti, D.S.; Cheatham, T.; Darden, T.; Duke, R.; Giese, T.J.; Gohlke, H.; Götz, A.; Greene, D.; Homeyer, N.; Izadi, S.; Kovalenko, A.; Lee, T.-S.; LeGrand, S.; Li, P.; Lin, C.; Liu, J.; Luchko, T.; Luo, R.; Kollman, P.A., Amber 2017; University of California: San Francisco, United States of America, 2017, 10.13140/RG.2.2.36172.41606.

(10) Hawkins, P. C.; Skillman, A. G.; Warren, G. L.; Ellingson, B. A.; Stahl, M. T., Conformer Generation with OMEGA: Algorithm and Validation Using High Quality Structures from the Protein Databank and Cambridge Structural Database. *J. Chem. Inf. Model.* **2010**, *50*, 572-584.

(11) Liu, X.; Bai, F.; Ouyang, S.; Wang, X.; Li, H.; Jiang, H., Cyndi: A Multi-objective Evolution Algorithm Based Method for Bioactive Molecular Conformational Generation. *BMC Bioinf.* **2009**, *10*, 101.

(12) Forli, S.; Huey, R.; Pique, M. E.; Sanner, M. F.; Goodsell, D. S.; Olson, A. J., Computational Protein–ligand Docking and Virtual Drug Screening with the AutoDock Suite. *Nat. Protoc.* **2016**, *11*, 905.

(13) Jmol, an Open-source Java Viewer for Chemical Structures in 3D. Jmol Web Page: http://www.jmol.org/.

(14) Stierand, K.; Rarey, M., PoseView - Molecular Interaction Patterns at a Glance. *J. Cheminf.* **2010**, *2*, P50.

(15) Salentin, S.; Schreiber, S.; Haupt, V. J.; Adasme, M. F.; Schroeder, M., PLIP: Fully Automated Protein–ligand Interaction Profiler. *Nucleic Acids Res.* **2015**, *43*, W443-W447.

(16) Jiménez-Sánchez, C.; Lozano-Sánchez, J.; Gabaldón-Hernández, J. A.; Segura-Carretero, A.; Fernández-Gutiérrez, A., RP-HPLC–ESI–QTOF/MS2 Based Strategy for the Comprehensive Metabolite Profiling of Sclerocarya Birrea (Marula) Bark. *Ind. Crops Prod.* **2015**, *71*, 214-234.

(17) Lozano-Sanchez, J.; Segura-Carretero, A.; Menendez, J. A.; Oliveras-Ferraros, C.; Cerretani, L.; Fernandez-Gutierrez, A., Prediction of Extra Virgin Olive Oil Varieties Through Their Phenolic Profile. Potential Cytotoxic Activity Against Human Breast Cancer Cells. *J. Agric. Food Chem.* **2010**, *58*, 9942−9955.

(18) Santos, S. A.; Vilela, C.; Freire, C. S.; Neto, C. P.; Silvestre, A. J., Ultra-high Performance Liquid Chromatography Coupled to Mass Spectrometry Applied to the Identification of Valuable Phenolic Compounds from Eucalyptus Wood. *J. Chromatogr. B* **2013**, *938*, 65-74.

(19) Wang, H.; Yan, G.; Zhang, A.; Li, Y.; Wang, Y.; Sun, H.; Wu, X.; Wang, X., Rapid Discovery and Global Characterization of Chemical Constituents and Rats Metabolites of Phellodendri Amurensis Cortex by Ultra-performance Liquid Chromatography-electrospray Ionization/Quadrupole-time-of-flight Mass Spectrometry Coupled with Pattern Recognition Approach. *Analyst* **2013**, *138*, 3303-3312.

(20) Salucci, M.; Stivala, L.; Maiani, G.; Bugianesi, R.; Vannini, V., Flavonoids Uptake and Their Effect on Cell Cycle of Human Colon Adenocarcinoma Cells (Caco2). *Br. J. Cancer* **2002**, *86*, 1645-1651.

(21) Stagos, D.; Kazantzoglou, G.; Magiatis, P.; Mitaku, S.; Anagnostopoulos, K.; Kouretas, D., Effects of Plant Phenolics and Grape Extracts From Greek Varieties of Vitis Vinifera on Mitomycin C and Topoisomerase I-induced Nicking of DNA. *Int. J. Mol. Med.* **2005**, *15*, 1013-1022.

(22) Stulnig, T.; Waldhäusl, W., 11β-Hydroxysteroid Dehydrogenase Type 1 in Obesity and Type 2 Diabetes. *Diabetologia* **2004**, *47*, 1-11.

(23) Yabe-Nishimura, C., Aldose Reductase in Glucose Toxicity: A Potential Target for the Prevention of Diabetic Complications. *Pharmacol. Rev.* **1998**, *50*, 21-34.

(24) Aronoff, S. L.; Berkowitz, K.; Shreiner, B.; Want, L., Glucose Metabolism and Regulation: Beyond Insulin and Glucagon. *Diabetes Spectr.* **2004**, *17*, 183-190.

(25) Wagner, R.; Kaiser, G.; Gerst, F.; Christiansen, E.; Due-Hansen, M. E.; Grundmann, M.; Machicao, F.; Peter, A.; Kostenis, E.; Ulven, T., Reevaluation of Fatty Acid Receptor 1 as a Drug Target for the Stimulation of Insulin Secretion in Humans. *Diabetes* **2013**, *62*, 2106-2111.

(26) van Poelje, P. D.; Dang, Q.; Erion, M. D., Fructose-1, 6-bisphosphatase as a Therapeutic Target for Type 2 Diabetes. *Drug Discovery Today: Ther. Strategies* **2007**, *4*, 103-109.

(27) Saltiel, A. R.; Kahn, C. R., Insulin Signalling and the Regulation of Glucose and Lipid Metabolism. *Nature* **2001**, *414*, 799.

(28) Etxeberria, U.; de la Garza, A. L.; Campión, J.; Martinez, J. A.; Milagro, F. I., Antidiabetic Effects of Natural Plant Extracts Via Inhibition of Carbohydrate Hydrolysis Enzymes with Emphasis on Pancreatic Alpha Amylase. *Expert Opin. Ther. Targets* **2012**, *16*, 269-297.

(29) Martin, J.; Veluraja, K.; Ross, K.; Johnson, L.; Fleet, G.; Ramsden, N.; Bruce, I.; Orchard, M.; Oikonomakos, N., Glucose Analog Inhibitors of Glycogen Phosphorylase: The Design of Potential Drugs for Diabetes. *Biochemistry* **1991**, *30*, 10101-10116.

(30) Mellado-Gil, J. M.; Cobo-Vuilleumier, N.; Gauthier, B. R., Islet B-cell Mass Preservation and Regeneration in Diabetes Mellitus: Four Factors with Potential Therapeutic Interest. *J. Transplant.* **2012**, *2012*, 230870.

(31) Monsalve, F. A.; Pyarasani, R. D.; Delgado-Lopez, F.; Moore-Carrasco, R., Peroxisome Proliferator-activated Receptor Targets for the Treatment of Metabolic Diseases. *Mediators Inflammation* **2013**, *2013*, 549627.

(32) Jeoung, N. H., Pyruvate Dehydrogenase Kinases: Therapeutic Targets for Diabetes and Cancers. *Diabetes Metab. J.* **2015**, *39*, 188-197.

(33) Herman, M. A.; Kahn, B. B., Glucose Transport and Sensing in the Maintenance of Glucose Homeostasis and Metabolic Harmony. *J. Clin. Invest.* **2006**, *116*, 1767-1775.