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

Horacio Pérez-Sánchez¹,*, Helena den-Haan¹,², Jorge Peña-García¹, Jesús Lozano-Sánchez³,⁴, María Encarnación Martínez Moreno¹, Antonia Sánchez-Pérez¹, Andrés Muñoz¹, Pedro Ruiz-Espinosa⁵, Andreia S.P. Pereira⁶, Antigoni Katsikoudi⁷, José Antonio Gabaldón Hernández¹, Ivana Stojanović⁸, Antonio Segura Carretero³,⁴,* and Andreas G. Tzakos⁷,*

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.
**ABSTRACT:** The DIA-DB is a web server for the prediction of diabetes drugs that uses two different and complementary approaches; a) comparison by shape similarity against a curated database of approved anti-diabetic drugs and experimental small molecules, and b) inverse virtual screening of the input molecules chosen by the users against a set of therapeutic protein targets identified as key elements in diabetes. As a proof of concept DIA-DB was successfully applied in an integral workflow for the identification of the anti-diabetic chemical profile in a complex crude plant extract. To this end, we conducted the extraction and LC-MS based chemical profile analysis of *Sclerocarya birrea*, and subsequently utilized this data as input for our server. The server is open to all users, registration is not necessary, and a detailed report with the results of the prediction is sent to the user by E-mail once calculations are completed. This is a novel public domain database and web server specific for diabetes drugs and can be accessed online through http://bio-hpc.eu/software/dia-db/.

**1. INTRODUCTION**

Diabetes mellitus was the direct cause of death for roughly 1.5 million people in 2012 and it is estimated that it will be the 7th leading cause of death in 2030\(^1\). There are an estimated 300 million diabetes patients in the world\(^2\) but considering that only 50% of people afflicted with diabetes are actually diagnosed, this number may be closer to 600 million diabetes patients worldwide. In a recent report by the International Diabetes Federation, the global landscape is expected to rise to 592 million by 2035\(^3\).

The basic aspect of diabetes is the elevated levels of blood sugar caused by the inability of the cells to absorb glucose, either because of low insulin levels (type 1 diabetes – T1D)\(^4\) or insulin resistance (type 2 diabetes – T2D)\(^5\). T1D is characterized as an autoimmune disorder where the cells of the immune system attack and destroy the pancreatic \(\beta\)-cells resulting in decreased insulin secretion\(^6\)–
T2D develops as a consequence of resistance of peripheral tissues such as liver, muscles and adipose tissue to insulin stimulation (insulin resistance). The pathophysiological processes leading to T2D include deterioration of β-cells functions, chronic hyperglycemia, and insulin resistance in musculoskeletal and adipose tissues.

Current medical care employs a wide array of pharmacological and lifestyle interventions aimed at managing hyperglycemia. Although it has been evidenced that, following both suitable diet and physical exercise the incidence of diabetes is decreased by up to 58%, the middle-term dropout rate is greater than 65%. Common anti-diabetic drugs available to treat patients with T2D include sulfonylureas (i.e., Gliclazide), thiazolidinediones (i.e., Pioglitazone), biguanides (i.e., Metformin), dipeptidyl peptidase-4 inhibitors (i.e., Sitagliptin), glinides (i.e., Repaglinide) and α-glucosidase inhibitors (i.e., Acarbose). However, these can become ineffective over time. In addition, none of these anti-diabetic drugs are free from adverse effects including the risk of hypoglycemia, gastrointestinal disturbances, weight gain, diarrhea, renal failure and hypersensitivity. Thus, there is a need for new and, if possible, more effective drugs and to make use of prevention strategies due to the premature morbidity and mortality associated with the disease that could potentially burden personal, as well as annual national, healthcare expenditures.

Because of the extent and severity of diabetes, anti-diabetic research has been developing rapidly over the years and aims to fully understand the biochemistry of the disease and to develop cheaper and more efficient drugs. Although numerous studies have been made over the last 20 years, including clinical trials and drug discoveries (mostly for T2D) it is rather challenging and time-consuming for researchers to assemble all the new information about that progress, since it is scattered throughout the web. It can be clearly seen that there is an unmet need for a free, easily accessible and handy online server that will not only operate as a database for information about
every anti-diabetic drug and promising experimental compound but also as a guide for drug design and development. This prompted us to develop the DIA-DB, a pioneering online server that, on the one hand, integrates and presents all the information about the activity of known anti-diabetic compounds and, on the other hand, can be an effective tool for the drug design, development and re-purposing of drugs against diabetes.

This is an expansion of our former work\textsuperscript{12}. In this new and vastly improved DIA-DB version, the database has been enhanced and a new method for virtual screening has been implemented. In addition, new tools for 3D visualization of the results have been integrated. We also illustrate the capacity and validity of our tool towards the prediction and experimental verification of the anti-diabetic potential of \textit{Sclerocarya birrea}. Along these lines we conducted the extraction and LC-MS based chemical profile analysis of a \textit{Sclerocarya birrea} stem-bark aqueous extract, and subsequently utilized this data as input for our server. DIA-DB pinpointed that the anti-diabetic potential of \textit{Sclerocarya birrea} stem-bark extract may be the result of the collective action of multiple bioactive compounds regulating and restoring several dysregulated interconnected diabetic biological processes.

\section*{2. MATERIALS AND METHODS}

\textbf{2.1. DIA-DB database design.} The DIA-DB is a database composed of two related tables namely Drugs and Calculations. The Drugs table stores information about anti-diabetic compounds: name, structure, SMILES (simplified molecular-input line-entry system), status (approved/experimental), etc (Table S1). Note that the information in the Drugs table is shared with the Calculations tables, as it contains the core DIA-DB data. The Calculations tables stores information about docking experiments and structure similarity comparison requests. The docking and similarity comparison calculations, utilize the anti-diabetic compound data that are included
in the Drugs table. For the similarity-based calculations, the global three-dimensional shape similarity search can be performed with WEGA\textsuperscript{13} or SHAFTS\textsuperscript{14} tools. Autodock Vina is used to perform the docking-based calculations\textsuperscript{15}. A partial view on the relational models containing the most relevant tables is shown in Figure 1 for the Drugs and Calculations tables. In the first case, the tables for representing anti-diabetic drug data are mainly “Compounds”, “Compound family”, “Ligand” and “Protein target”. Note that the compound family, ligand and protein targets are considered as proper entities which are related to the compound’s core data.

Regarding the tables representing the Calculations data, the most relevant ones are “Experiments”, including its status and start/end dates; “Users”, containing contact data (E-mails) relevant to DIA-DB users; and “Similarity” and “Result Experiment” for structure similarity comparisons and docking experiments, respectively. Note that the “Similarity” table has a relationship with the “Compounds” from the Drug Table as it needs to relate the source and target ligands.

DIA-DB has been deployed on an Apache 2 server running Ubuntu 14.04.5 64-bits. User interfaces follow a friendly and interactive web design obtained by the combination of several technologies such as JavaScript\textsuperscript{16}, Jquery\textsuperscript{17}, PHP and HTML. Data managed by DIA-DB is stored in a relational database through the MySQL database management system\textsuperscript{18}. The content of this database is continuously maintained, and newly discovered compounds will be added. All docking and shape similarity queries are sent to a remote computer cluster where a SLURM\textsuperscript{19} based queue manager distributes the jobs. Results are sent to the user by an E-mail with a link to a web page where they are displayed in HTML format.

The system is available at http://bio-hpc.eu/dia-db, and registration is not necessary. A set of detailed tutorials, examples, frequently asked questions and related publications, among others, are offered.
Figure 1. (a) Relational model for the Drugs Table (partial) and (b) relational model for the Calculations table (partial). Note that the “Compounds” from the Drugs table is reused for the experiment data.
There are several ways to perform queries in DIA-DB: searching by compound name or SMILES (Drugs table), or by submitting a docking study (Calculations table). In the first case, a detailed view of the resulting compounds can be obtained online, in real time (Figure S1). In the second case, for docking studies, a link for the detailed report with the prediction docking results are sent to the user by E-mail once simulations are completed (Figure S2). This task can take from several hours, to two or three days, depending on the query compound and cluster occupation. A detailed description of the methodology implemented in the design of the Drugs and Calculations Tables as well as a detailed description of the experimental workflow is available as Supporting Information.

2.2. Case study – characterization of *Sclerocarya birrea*. A detailed description of extraction and analytical characterization of the *Sclerocarya birrea* stem-bark aqueous extract is available as Supporting Information.

3. RESULTS AND DISCUSSION

3.1 The DIA-DB database. In an effort to navigate the chemical space related to diabetes, we created a database composed of current diabetes drugs and bioactive compounds of established anti-diabetic activity and also included their characterized protein targets. This also prompted us to develop, besides the database of drugs, tools that could allow navigation on the anti-diabetes chemical space. We developed tools whereby this navigation can be conducted in two different ways namely a ligand-based approach and a structure-based approach. Through the ligand-based approach the compound of interest is screened against known anti-diabetes drugs and bioactive diabetes compounds (Figure S3).
This ligand-based similarity approach of the targeted compound (this could be a new synthetic molecule, a natural product or a known drug) with known diabetes drugs, could reveal unprecedented scaffold similarity that could imply common function and activity. This, since two isosteric molecules (retain similar volumes and shapes) could be also considered bioisosteres given the probability to recognize a similar or the same protein environment. This bioisosterism in chemical molecules could provide the predictive capacity power to determine for a given compound potent anti-diabetic activity. To provide flexibility to the user, the specific tool can be used through submission of SMILES of the desired compound or via drawing a structure of the compounds in an appropriate drawing window and submit for the performance of similarity metrics. Moreover, a tool that can directly add the name of the compound and directly search for a compound is offered. This database was created and curated manually for compounds with known anti-diabetic properties.

Where the ligand-based similarity approach implies function and activity, with the structure-based docking approach the user can determine whether a given compound may actually interact with the active binding site of the protein and form the necessary interactions (hydrogen bonds, Van der Waals interactions, ionic bonds, π-π bonds) needed to bring about the desired effect of inhibition/activation. Also, the structure-based docking approach may identify new amino acids important for ligand binding that may significantly potentiate activity. Eighteen protein targets associated with diabetes were selected for the structure-based docking approach (Table S3). The major anti-diabetic effects commonly studied in vivo are the effect of a given compound or plant extract on hyperglycaemia, hyperlipidaemia and plasma insulin levels. The 18 protein targets of the docking-based approach in the DIA-DB can be divided into three categories representing either their mode of action on insulin secretion and/or sensitivity (DPP4, FFAR1, HSD11B1, INSR,
PTPN9, RBP4), regulation of glucose metabolism (AKR1B1, AMY2A, FBP1, GCK, MGAM, PDK2, PYGL) or regulation of lipid metabolism (NR5A2, PPARA, PPARD, PPARG, RXRA). By evaluating the potential effects of compounds on these targets, one may identify the anti-diabetic mechanisms of action of a given compound or plant extract. To investigate the applicability of our server we screened the anti-diabetic activity of several compounds identified from an aqueous extract of the stem-bark of *Sclerocarya birrea*.

### 3.2. Case study – characterization and target fishing with *Sclerocarya birrea*.

Several drugs are available to treat patients with T2D, however, these can become ineffective over time and have adverse effects. There is thus a continued effort to discover and design new and, if possible, more effective drugs to treat diabetes. One of the potential approaches would be focused on the survey of vast reserves of phytotherapeutics for diabetes prevention or treatment. Nowadays, the acceptance and use of natural products as alternative therapies are dramatically increasing. Driven by the desperate situation of diabetics, different opportunist companies have found a market niche. Currently there are on the market numerous alternative remedies and natural supplements that promise and ensure a cure for T2D.

Among them lies *Sclerocarya birrea*, also called Marula or elephant tree, used to manage various diseases including diabetes. Treatment of streptozotocin-induced diabetic rats with *Sclerocarya birrea* extracts is associated with a reduction in hyperglycaemia and improvement of oral glucose tolerance test, reduction in hyperlipidaemia and an increase in plasma insulin levels. The bioactive compounds and anti-diabetic mechanisms of action, however, remain largely unknown.

According to WHO requirements, the assessment of the heath promoting potential of a plant extract must be based on a valid scientific hypothesis and realistic studies supporting the hypothesis. In this sense, the strict analytical profile of a plant extract must be monitored and
controlled. In the case of *Sclerocarya birrea*, however, only general compositional analysis, now outdated, have been reported, revealing that these barks may contain chemical compounds such as gallotannins, flavonoids, alkaloids, steroids (including β-sitosterol), coumarins, triterpenoids, sesquiterpene hydrocarbons, ascorbic acid, oleic, myristic, stearic acids, and amino acids\textsuperscript{24}. Still, a large amount of chemical information remains unknown and needs to be identified before their relation to health can be fully understood.

The major type of compounds present in the extract were the flavan-3-ols in the form of (epi)catechin, (epi)gallocatechin, (epi)gallocatechin gallate monomers and dimers. A detailed description of the characterization of the compounds identified in the *Sclerocarya birrea* extract as well as a representative chromatogram (Figure S4) and tentative identification (Table S4) is available as Supporting Information. In this study the flavan-3-ols were identified to be potential regulators of DPP4, AKR1B1, HSD11B1, AMY2A, MGAM, PPARG, PPARD, AMPK and GCK (Figures 2 and S5). Several *in vitro* studies can be found in literature supporting and validating some of the targets identified here for the flavan-3-ols\textsuperscript{25-36}. 
The *in vivo* anti-diabetic effects observed following treatment with *Sclerocarya birrea* extracts may be the result of several of the compounds identified here regulating multiple protein targets whose biological roles are interconnected. The observed reduction in hyperglycemia and improvement in the oral glucose tolerance test may be attributed to regulation of GCK, AMPK, MGAM and AMY2A by the *Sclerocarya birrea* compounds. Inhibition of AMY2A and MGAM will delay carbohydrate digestion and thus lower the postprandial blood glucose level. Inhibitory activity of *Sclerocarya birrea* extracts of AMY2A and MGAM was observed *in vitro* in the study of Da Costa Mousinho *et al.*, 2013 and here we have identified gallic acid, citric acid, catechin, epicatechin, (epi)gallocatechin-3’-O-gallate, (epi)gallocatechin-3’-O-gallate, (epi)catechin di-O-gallate, procyanidin B2 and theaflavin 3,3’-digallate as being responsible for this activity. Activation of GCK will also lead to a reduction in serum glucose levels by promoting glycogenesis and glycolysis through the phosphorylation of glucose to glucose-6-phosphate. GCK can also act...
as a glucose sensor in pancreatic B-cells and stimulate insulin secretion\textsuperscript{39} and thus the observed \textit{in vitro} glucose-stimulated insulin secretion from INS-1E cells\textsuperscript{22} may be a result of GCK activation by the \textit{Sclerocarya birrea} compounds procyanidin B2, (epi)catechin-(epi)catechin-3’-O-gallate, trigalloyl shikimic acid, (epi)catechin di-O-gallate, (epi)afzelechin-O-gallate, nothofagin and theaflavin 3,3’-digallate. Activation of AMPK is associated with increased glucose uptake by the muscles and decreased glucose production by the liver resulting in a decrease in blood glucose levels\textsuperscript{33}. The observed \textit{in vitro} stimulation of glucose uptake by muscle, liver and adipose cells by \textit{Sclerocarya birrea} extracts\textsuperscript{38,40} may thus be the result of AMPK activation by compounds catechin, epicatechin, epigallocatechin, gallic acid and citric acid.

The targeting of proteins that promote insulin secretion and improve insulin sensitivity will in turn also promote glucose homeostasis and reduce hyperglycaemia. Inhibition of DPP4 will increase the half-life of the incretin hormones thereby increasing insulin secretion and allowing time to normalize blood glucose levels\textsuperscript{41}. Similarly, compounds capable of inhibiting HSD11B1 can inhibit glucose production by the liver and improve glucose-dependent insulin sensitivity\textsuperscript{42}. The compounds procyanidin B2, (epi)gallocatechin-3’-O-gallate-(epi)gallocatechin-3’-O-gallate, trigalloyl shikimic acid, (epi)catechin di-O-gallate, (epi)afzelechin-O-gallate and nothofagin were identified as potential inhibitors of both DPP4 and HSD11B1. Treatment of streptozotocin-induced diabetic rats was also associated with a reduction in hyperlipidaemia. The compounds catechin, epicatechin, (epi)gallocatechin gallate, gallic acid, (epi)catechin gallate, (epi)gallocatechin, (epi)afzelechin gallate, trigalloyl shikimic acid, (epi)catechin di-O-gallate and theaflavin 3,3’-digallate were found to be potential regulators of PPARD, PPARG and AMPK that play various roles in lipid metabolism regulation\textsuperscript{31,43,44}. 
4. CONCLUSION

In this paper, we describe the development of a novel chemoinformatic server the DIA-DB that integrates information on drugs and proteins implicated in diabetes. Through this information we developed different tools that a user can utilize to assist drug design for diabetes drugs. The user can either use a ligand-based approach or a structure-based approach to screen for hit target compounds. We envisage that our server will be of importance in the navigation of the chemical space towards the development of new diabetes drugs or repurposing of existing drugs against diabetes. To pinpoint the predictive capacity of the DIA-DB we conducted an integral workflow for *Sclerocarya birrea* starting from the extraction and LC-MS based chemical profile analysis and then utilized this recorded data as input for our server. Several studies in literature could be found supporting and validating some of the protein targets identified by the DIA-DB for the flavan-3-ols. The docking- and similarity-based results indicated that the anti-diabetic potential of *Sclerocarya birrea* stem-bark extracts may be the result of the collective action of multiple bioactive compound regulating and restoring several dysregulated interconnected diabetic biological processes.

ACKNOWLEDGEMENTS

This work has been funded by grants from the Spanish Ministry of Economy and Competitiveness (CTQ2017-87974-R and TIN2016-78799-P), by Fundación Séneca la Región de Murcia under Project 20988/PI/18, and by Herbafor SL under a grant from Instituto de Fomento de la Región de Murcia under project 2015.08.ID+1.0072. This research was partially supported by the supercomputing infrastructure of Poznan Supercomputing Center, the e-infrastructure program of the Research Council of Norway via the supercomputer center of UiT - the Arctic University of Norway, and by the supercomputing infrastructure of the NLHPC (ECM-02), Powered@NLHPC.
This research project has been co-financed by the European Union (European Regional Development Fund – ERDF) and Greek national funds through the Operational Program ‘‘THESSALY-MAINLAND GREECE AND EPIRUS-2007–2013’’ of the National Strategic Reference Framework (NSRF 2007–2013).

CONFLICT OF INTEREST

All authors declare no conflict of interest

REFERENCES


