Article

Indexing Natural Products for Their Potential Anti-Diabetic Activity: Filtering and Mapping Discriminative Physicochemical Properties

Mouhammad Zeidan [1,†], Mahmoud Rayan [2,†], Nuha Zeidan [3,†], Mizied Falah [4,5,†] and Anwar Rayan [2,6,*]

[1] Molecular Genetics and Virology Laboratory, QRC-Qasemi Research Center, Al-Qasemi Academic College, P.O. Box 124, Baka EL-Garbiah 30100, Israel; mouhammad.zeidan7@gmail.com

[2] Institute of Applied Research-Galilee Society, P.O. Box 437, Shefa-Amr 20200, Israel; mahmoud_ryan@hotmail.com

[3] Clalit Health Service, Diet and Nutrition Unit, P.O. Box 789, Arara 30026, Israel; nuhayz@gmail.com

[4] Eliachar Research Laboratory, Galilee Medical Center, P.O. Box 21, Nahariya 22100, Israel; MiziedF@gmc.gov.il

[5] Faculty of Medicine in the Galilee, Bar-Ilan University, Ramat Gan 52900, Israel

[6] Drug Discovery Informatics Laboratory, QRC-Qasemi Research Center, Al-Qasemi Academic College, P.O. Box 124, Baka EL-Garbiah 30100, Israel

* Correspondence: a_rayan@qsm.ac.il; Tel.: +972-4628-6764

† These authors contributed equally to this work.

Abstract: Diabetes mellitus (DM) poses a major health problem, for which there is an unmet need to develop novel drugs. The application of in silico techniques and optimization algorithms is instrumental to achieving this goal. A set of 97 approved anti-diabetic drugs, representing the active domain, and a set of 2892 natural products, representing the inactive domain, were used to construct predictive models and to index anti-diabetic bioactivity. Our recently-developed approach of 'iterative stochastic elimination' was utilized. This article describes a highly discriminative and robust model, with an area under the curve above 0.96. Using the indexing model and a mix ratio of 1:1000 (active/inactive), 65% of the anti-diabetic drugs in the sample were captured in the top 1% of the screened compounds, compared to 1% in the random model. Some of the natural products that scored highly as potential anti-diabetic drug candidates are disclosed. One of those natural products is caffeine, which is noted in the scientific literature as having the capability to decrease blood glucose levels. The other nine phytochemicals await evaluation in a wet lab for their anti-diabetic activity. The indexing model proposed herein is useful for the virtual screening of large chemical databases and for the construction of anti-diabetes focused libraries.

Keywords: diabetes mellitus; anti-diabetic drugs; drugs analysis; ligand-based screening approach; bioactivity index

1. Introduction

Diabetes mellitus, an expanding pandemic worldwide, is expected [1,2] to affect more than 640 million individuals by 2040 [3]. Diabetes mellitus type 2 (T2DM), one of three types of diabetes [1,4], accounts for more than 90% of all cases now diagnosed at any age, even in children [5]. T2DM is characterized by gradually progressive insulin resistance in various body tissues (liver, muscle and adipose), or failure in islet β-cells, or both [6–8], leading to the development of chronic hyperglycemia [9,10]. In T2DM-affected individuals, other cardiovascular risk factors (hypertension, dyslipidemia and obesity) are abundantly present [4,11–13]. If uncontrolled, T2DM leads to

nephropathy, neuropathy, retinopathy, and amputations and poses a high risk of cardiovascular and cerebro-vascular events [4,14–18]. Due to the multifactorial nature of T2DM, no single anti-hyperglycemic agent can correct all abnormalities [10]. At this point in time, glucose control is the major focus of the management of T2DM, along with the management of cardiovascular risk factors, which includes reducing the duration of the disease, smoking cessation, the adoption of healthy lifestyle habits, blood pressure control, lipid management, patient adherence and resources and, in some cases, antiplatelet therapy [19–30]. The complexities of treatment entail enormous health, economic and social burdens. In practice, soon after diagnosis, metformin 'monotherapy' or 'dual therapy' is initiated, consisting of one of six treatment options: (1) sulfonylurea; (2) thiazolidinediones (TZD); (3) dipeptidyl peptidase 4 inhibitors (DPP-4 inhibitor); (4) selective sodium glucose cotransporter 2 inhibitor (SGLT2 inhibitor); (5) glucagon-like peptide 1 receptor agonists (GLP-1 receptor agonist); or (6) basal insulin. When the glycated hemoglobin (HbA_{1c}) target is not achieved, 'triple therapy' combinations that do not include metformin may also be considered [19]. However, there is a limited potential with the use of any of the available drugs in patients with T2DM because balancing the glucose-lowering efficacy, the side-effect profiles, the anticipation of additional benefits, cost and other practical aspects of care is difficult. Moreover, data on the side effects of most of the possible drug combinations are lacking. Hence, the quest for new compounds that can complement current therapies for T2DM and its comorbidities is expanding [31,32]. Notably, natural products offer many potential mechanisms of actions for improving glucose homeostasis, which could reduce and/or abolish diabetic complications [32–34]. Since drug discovery and development involve time-consuming and expensive processes, computer methodologies are utilized to shorten the time span of drug development and to reduce costs, and in silico techniques may contribute to the identification of new lead compounds and to the optimization of drugs in clinical use [35].

In efforts to detect novel bioactive ligands, ligand-based techniques, including properties-based and pharmacophore-based tools, are being used more and more for modeling the bioactivity of molecules and for the virtual screening of large chemical databases [35–38]. Ligand-based modeling tools use optimization algorithms such as Monte Carlo simulations (MCs), simulated annealing (SA) [39], genetic algorithms (Gas) [40], neural networks (NNs) [41], support vector machines (SVM) [42], the k-nearest neighbor algorithm (kNN) [43,44], Bayesian classifiers and some combinations thereof (Monte Carlo/ simulated annealing algorithm, MCSA) [45–49]. Distinguishing between active and inactive ligands that are useful for treating a certain disease may be accomplished by using sets of active and inactive chemicals and certain optimization techniques [50–52]. Such techniques presume that bioactive ligands have common features that are not easily recognizable if only a small number of bioactive ligands are tested. Therefore, if a database contains a larger number of active/inactive ligands, drawing more significant and robust conclusions about the properties of these chemicals is achievable. In addition, it is essential to include in the training set of inactive chemicals, chemicals that cover the same property space as those in the screened chemical databases. In this way, analyses of sets of active/inactive ligands may shed light on characteristics leading to the bioactivity of active ligands. Due to the large number of descriptors taken into consideration during the modeling process, special optimization techniques that can overcome the limitations of the combinatorial nature of the problem of molecular bioactivity indexing are required. To address this, over the last decade, we have developed a new optimization algorithm capable of scanning multi-dimensional space and detecting the best solutions (finding the global minimum, as well as the set of local minima). Termed iterative stochastic elimination (ISE), it has been applied to solving structure-based problems [53], as well as ligand-based problems [54,55]. ISE uses an algorithm that can efficiently scan a multi-dimensional space in order to detect the best set of solutions. It has been used to solve problems such as positioning protons [56] and predicting side-chain conformations in proteins [57], scanning the conformational space of loops [58], searching the conformational space of cyclic peptides [59] and loops, predicting drug-likeness and indexing chemicals for their hERG liability [44,55].

Here, we report on the use of the ISE algorithm to construct a model for indexing natural products for their potential anti-diabetic activity and for mapping their discriminative physicochemical properties. Anti-diabetic drugs may act through numerous mechanisms of action and bind with different biological targets. However, as demonstrated by our previous experience [46], the proposed filters-based indexing approach can deal effectively with such complex problems. For this study, we have chosen to screen a database of natural products because they are secondary metabolites of organisms, which means they have been optimized to interact with biological systems through a long natural selection process [60–62], and thus, they serve as good drug candidates. It is worth noting that anti-diabetic drug candidates are molecules that, according to the model, have a high chance of exhibiting anti-diabetic activity, but they should be validated using in vivo methods in order to be considered truly anti-diabetic.

2. Results and Discussion

2.1. Utilization of the Iterative Stochastic Elimination Algorithm for Indexing Natural Compounds for Their Potential Anti-Diabetic Drug Likeness

In this study, we used the ISE algorithm to construct a model for indexing natural products for their potential anti-diabetic activity and to map their discriminative physicochemical properties. For the scanning, we used a set of 97 anti-diabetic drugs (presented in the simplified input line-entry system (SMILES) format in the Supporting Information (Table S1)) to represent the active domain. A database composed of 2892 natural products (which was prepared by collecting phytochemicals isolated from more than 800 different plants and which is distributed worldwide and available for purchase from AnalytiCon Discovery company, Postdam, Germany (www.ac-discovery.com)), sharing intermolecular similarity values <0.9, was used to represent the inactive domain. It is worth noting that the 2892 natural products are putative inactives and probably contain a few active chemicals. In order to guarantee that our active/inactive classes would not be biased by similar structures, we first checked for diversity among the 97 anti-diabetic drugs and the 2892 natural products and found them to be very diverse (see Figure 1A,B).
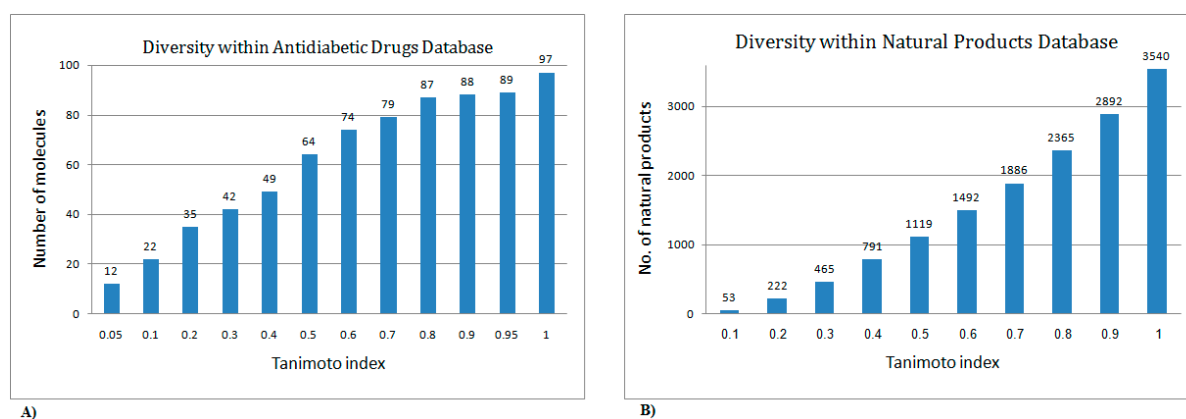

Figure 1. Diversity within anti-diabetic drugs (**A**) and diversity within the natural products database (**B**).

Optimal differentiation between the active and inactive chemicals was attained by searching in multivariable space for the best sets of descriptors ('variables') and the best ranges for all descriptors in each set, capable of distinguishing between actives and inactive chemicals. Since chemical descriptors generally interact with each other, changes in the range of one descriptor could have an effect on the best range of another descriptor, and the optimization process is obliged to take into consideration all descriptors in the set at one time to attain the best set of filters. A flowchart of the modeling process is

shown in Figure 2. For further details on applying ISE to obtain the best ranges for a set of descriptors and for the optimization process, see our previously-reported studies [46,55].
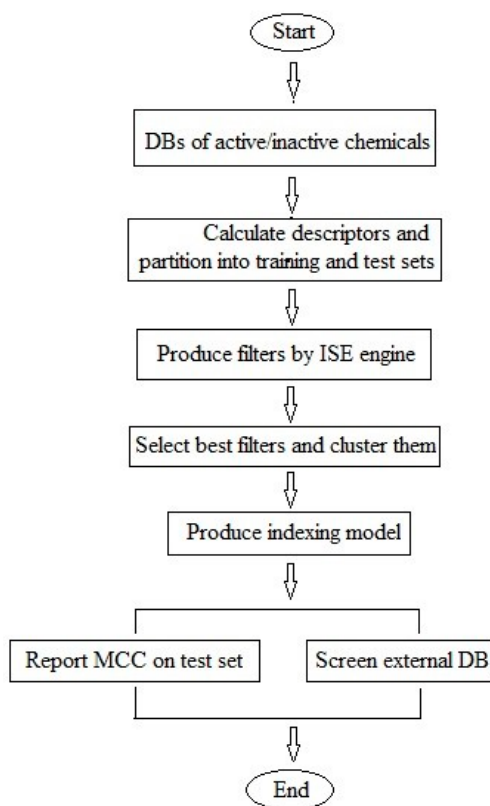

Figure 2. Flowchart for the ligand-based modeling process. ISE, iterative stochastic elimination; MCC, Matthews correlation coefficient.

2.2. Mapping the Discriminative Physicochemical Properties Responsible for Anti-Diabetic Activity

More than 65% of the anti-diabetic drugs had an intermolecular Tanimoto index of similarity <0.7. It is interesting to note that 89.7% of the anti-diabetic drugs obey Lipinski's rule of five (ROF) [63], and 74.2% obey Oprea's rule for lead-likeness [64] (see Figure 3).
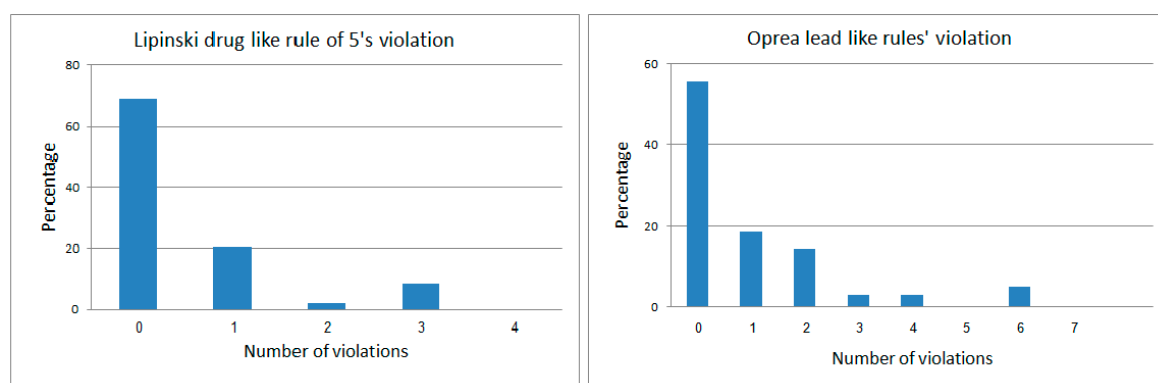

Figure 3. Violation distribution of anti-diabetic drugs for Lipinski's rule of drug-likeness and Oprea's rule for lead likeness.

Figure 4 shows the distribution plots of the Lipinski and Oprea physico-chemical properties of the set of anti-diabetic drugs.
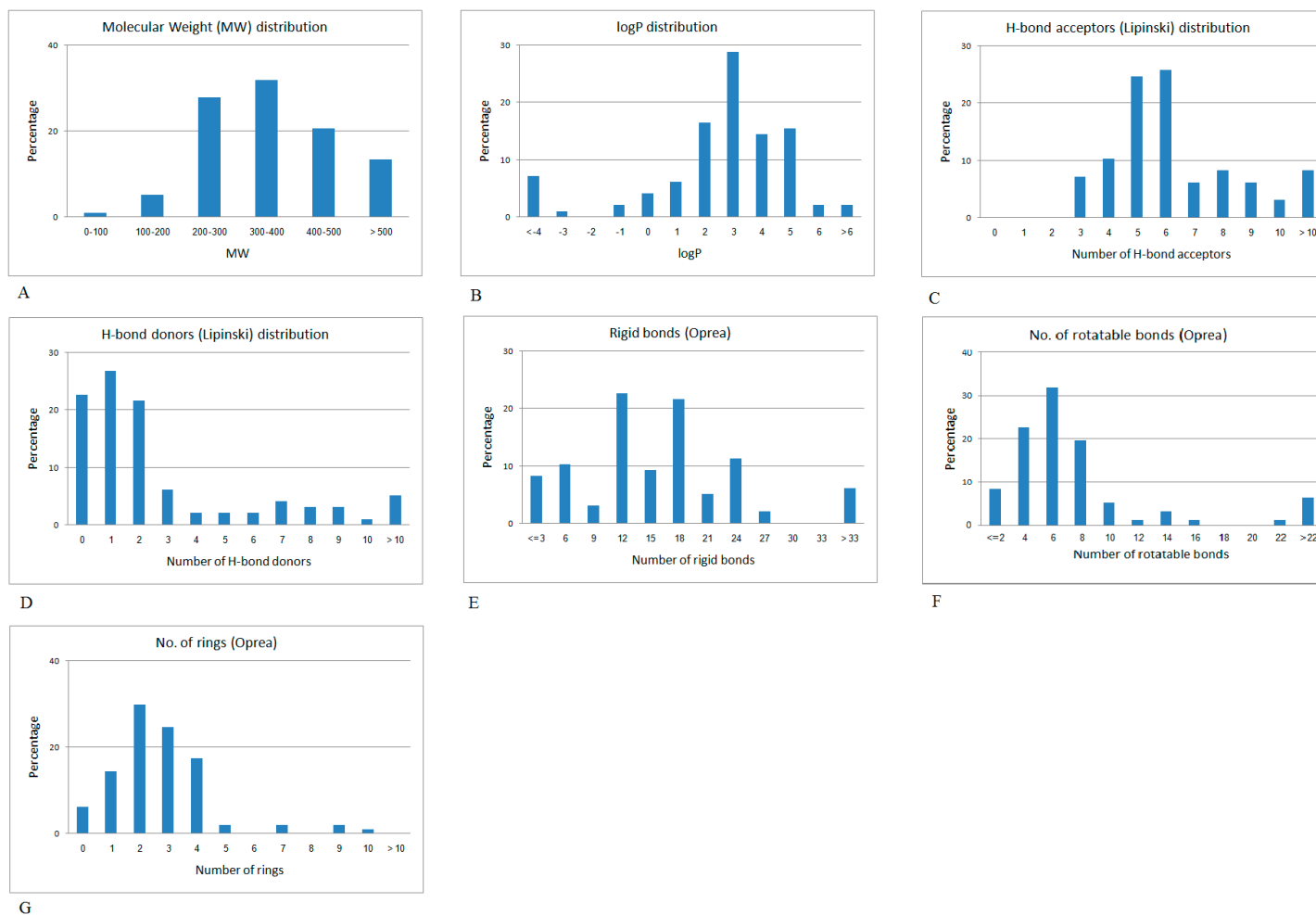
Figure 4. Physicochemical property distribution of anti-diabetic drugs. (**A**) Molecular weight distribution; (**B**) logP values; (**C**) the number of H-bond acceptors (lip_acc as coded by MOE software); (**D**) the number of H-bond donors (lip_don as coded by MOE software); (**E**) the number of rigid bonds; (**F**) the number of rotatable bonds; and (**G**) the number of aromatic atoms.

2.3. Filters and Descriptors Produced for Constructing the Predictive Model for Indexing Chemicals

To construct the indexing model, we used the ISE algorithm to produce 47 unique filters, composed of different sets and ranges of descriptors. As examples, three such filters are described in Table 1.

Table 1. Efficiency and descriptor ranges of three of the 47 filters used to produce the anti-diabetic activity indexing model.

Filter 1	Filter 2	Filter 3
MCC = 0.642	MCC = 0.640	MCC = 0.638
TP = 86.59%	TP = 62.88%	TP = 81.4%
TN = 77.42%	TN = 97.23%	TN = 82.3%
a_n O (0–6.99)	BCUT_PEOE_2 (0–0.67)	SMR_VSA3 (0–37.82)
PEOE_VSA + 4 (0–30.95)	GCUT_SLOGP_2 (0.11–0.27)	SMR_VSA1 (0–94.72)
Chiral (0–4.99)	Chiral_u (0–3)	GCUT_PEOE_3 (0–2.88)
SlogP_VSA2 (0–65.17)	GCUT_SLOGP_0 (−2.26−−0.91)	Reactive (0–0.00)

NOTE: The efficiency of the filters, in terms of their MCCs, is very close, but they differ in their true positive and negative percentages. In addition, the filters could be composed of different sets and/or ranges of descriptors. The name of descriptors stated herein are the same as named by CCG's computational suite MOE. A full description and methods of descriptors' calculation could be found in the site of Chemical Computing Group [65].

The efficiency of the three filters, in terms of their Matthews correlation coefficients, (MCCs), was very close, but they differed in their true positive and negative percentages, as well as in the index attached to each molecule. Filter 2 in Table 1 has an MCC of 0.64, and it identified successfully nearly 63% of the anti-diabetic drugs (true positives), while only 2.77% of the natural products database (presumably inactive) was misclassified (namely, turned up as false positives).

An analysis of the composition of the best filters disclosed a list of discriminative descriptors and/or physico-chemical properties. The data shown in Table 2 describe the most dominant descriptors in the 47 filters used to produce the anti-diabetic activity indexing model. The third column describes how many times the appearances of the descriptor compared to random.

Table 2. Number of appearances of descriptors within the set of the best unique filters. The full list of descriptors is presented in the Supporting Information (Table S2). Definition for descriptors and methods of its calculation could be found in the site if Chemical Computing Group [65].

Descriptor Name	No. of Appearances	Redundant More Times than Random
GCUT_SLOGP_0	24	23.7
a_ICM	16	15.8
PEOE_VSA + 4	12	11.9
SMR_VSA1	10	9.9
logS	9	8.9
Nmol	9	8.9
lip_druglike	9	8.9
Chi1_C	8	7.9
GCUT_PEOE_0	8	7.9
opr_leadlike	7	6.9
Q_VSA_FPOS	7	6.9
SMR_VSA3	7	6.9
a_don	6	5.9
a_hyd	6	5.9

Furthermore, Figure 5 was constructed by utilizing the WORDLE module (a tool for generating "word clouds" from text) and shows the number of appearances of descriptors in a graphic manner. For a full list of all of the descriptors used in the modeling process and their redundancy within the selected best filters, see the Supporting Information (Table S2) in the Appendix. The most dominant descriptors may be valuable for discriminating between anti-diabetic chemicals and inactive ones.
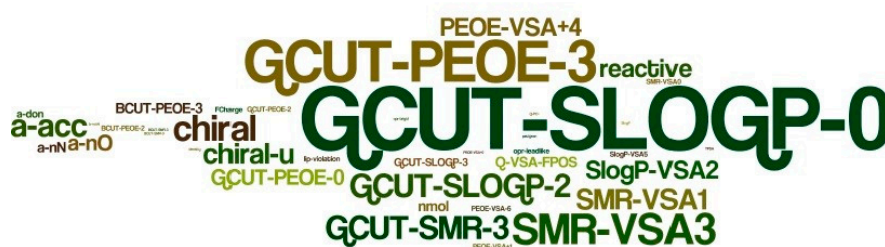
Figure 5. Appearances of descriptors in the 47 filters used to produce the anti-diabetic activity indexing model.

2.4. Assessing the Quality of Anti-Diabetic Activity-Indexing Model Potential

To assess the quality of the indexing model, based on the 47 range-based filters, various parameters such as the enrichment factor, Matthews's correlation coefficient (MCC), ROC curve and the area under the ROC curve (AUC) were generated. The percentage of true/false positives (left *y*-axis) and MCCs (right *y*-axis) were plotted against the molecular bioactivity index (MBI) threshold (*x*-axis), and the results are shown in Figure 6.
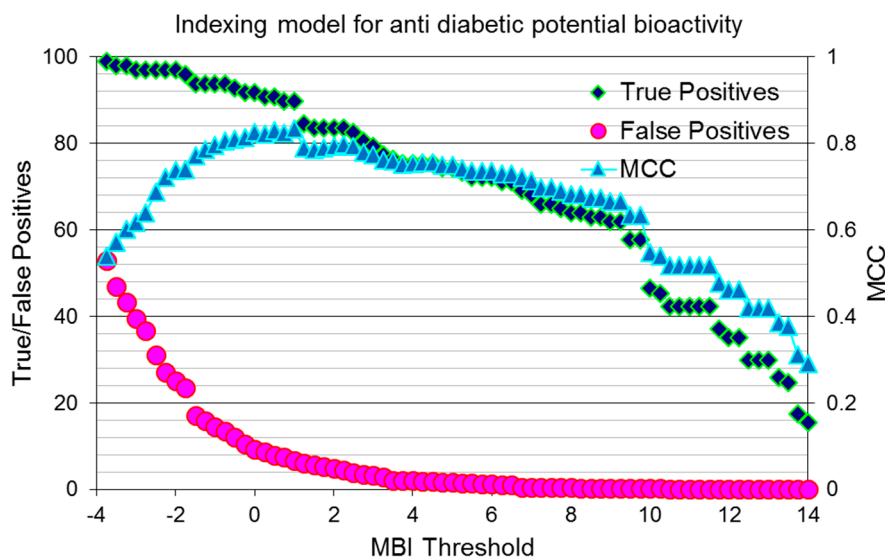

Figure 6. Indexing model for potential anti-diabetic activity: true/false positive percentage (left *y*-axis) and MCCs (right *y*-axis) plotted against the molecular bioactivity index (MBI) threshold (*x*-axis).

To illustrate how anti-diabetic drug candidates can be identified if natural products are sorted according to the model's predictions, rather than based on random selection, an enrichment plot was generated, and it is shown in Figure 7. An enrichment plot of the ISE-based model showing near-perfect results very close to the perfect model at the top fraction indicates a high prioritization power for the proposed model.
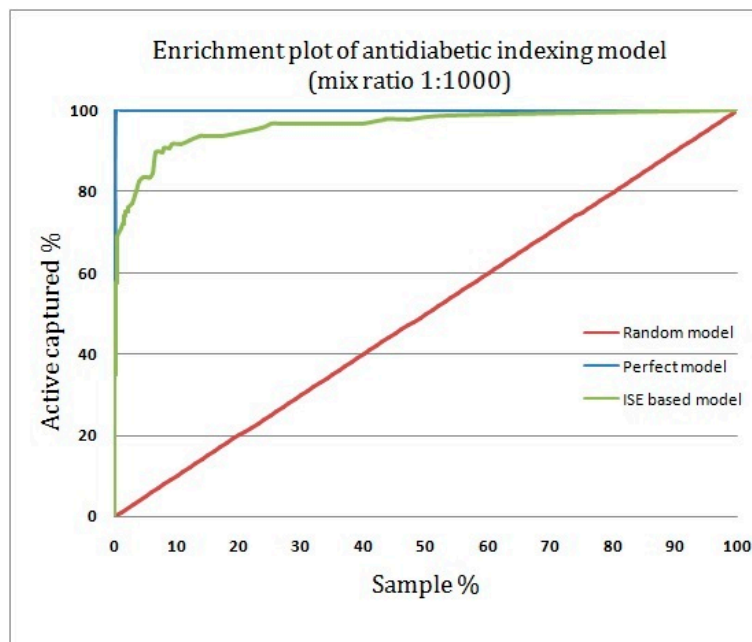
Figure 7. Enrichment plot of the anti-diabetic activity-indexing model.

With the use of the proposed anti-diabetic activity indexing model and a mix ratio of 1:1000 for active to inactive compounds, 65% of the anti-diabetic drugs were captured in the top 1% of the screened compounds, compared to 100% in the perfect model and 1% in the random model. This means that the enrichment factor at the top fraction of 1% is 65-fold (Figure 8).
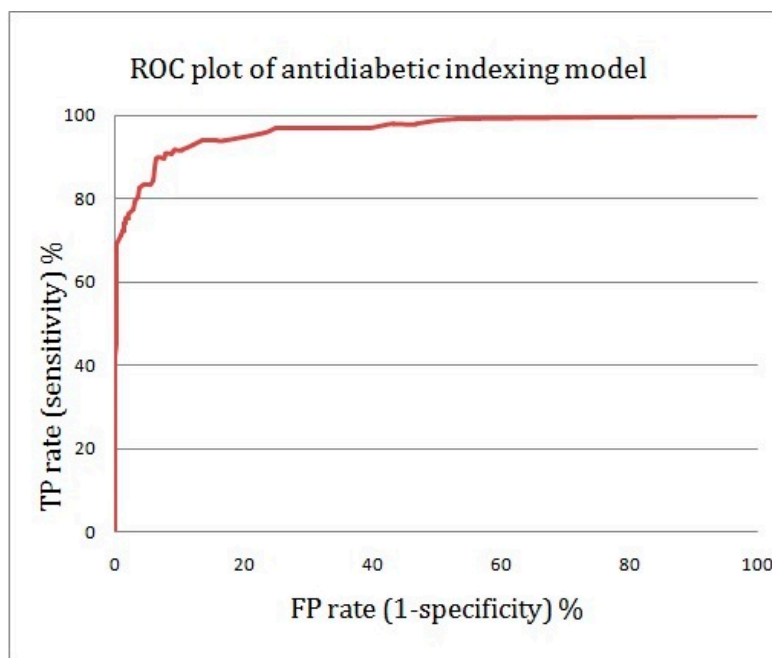

Figure 8. A receiver operating characteristic (ROC) curve showing the performance of the anti-diabetic activity-indexing model (the true positive rate is plotted against the false positive rate).

If we select molecules with an MBI above 13.0, the ISE-based model and the perfect model totally overlap. Thus, it seems that the proposed model is highly discriminative and exhibits very good

performance for the classification of anti-diabetic drug candidates versus inactive natural products. The area under the curve (AUC) attained was slightly above 0.96, which indicates that the model is very good.

2.5. New Potential Anti-Diabetic Drug Candidates as Disclosed by the ISE-Indexing Model

The database composed of 2892 natural products was virtually screened using the aforementioned filter-based indexing model. We assume that few chemicals in the database are anti-diabetic and will get a high MBI score. The MBI score, as shown in Figure 6, ranges between −4.0 (the lowest score) and 14.0 (the highest score). Figure 9 shows ten natural products that scored highly as potential anti-diabetic drug candidates according to our ISE-based anti-diabetic activity indexing model (with MBI score above 8.0). Using the threshold of MBI 8.0, the ratio of TP: FP is equal to 185:1. A search on PubMed revealed that one of the highly indexed phytochemicals (caffeine) has already been tested in vivo and found to be capable of decreasing blood glucose levels, with confirmed anti-diabetic activity [66–68]. The other nine phytochemicals await evaluation in the wet lab to ascertain their potential anti-diabetic activity.
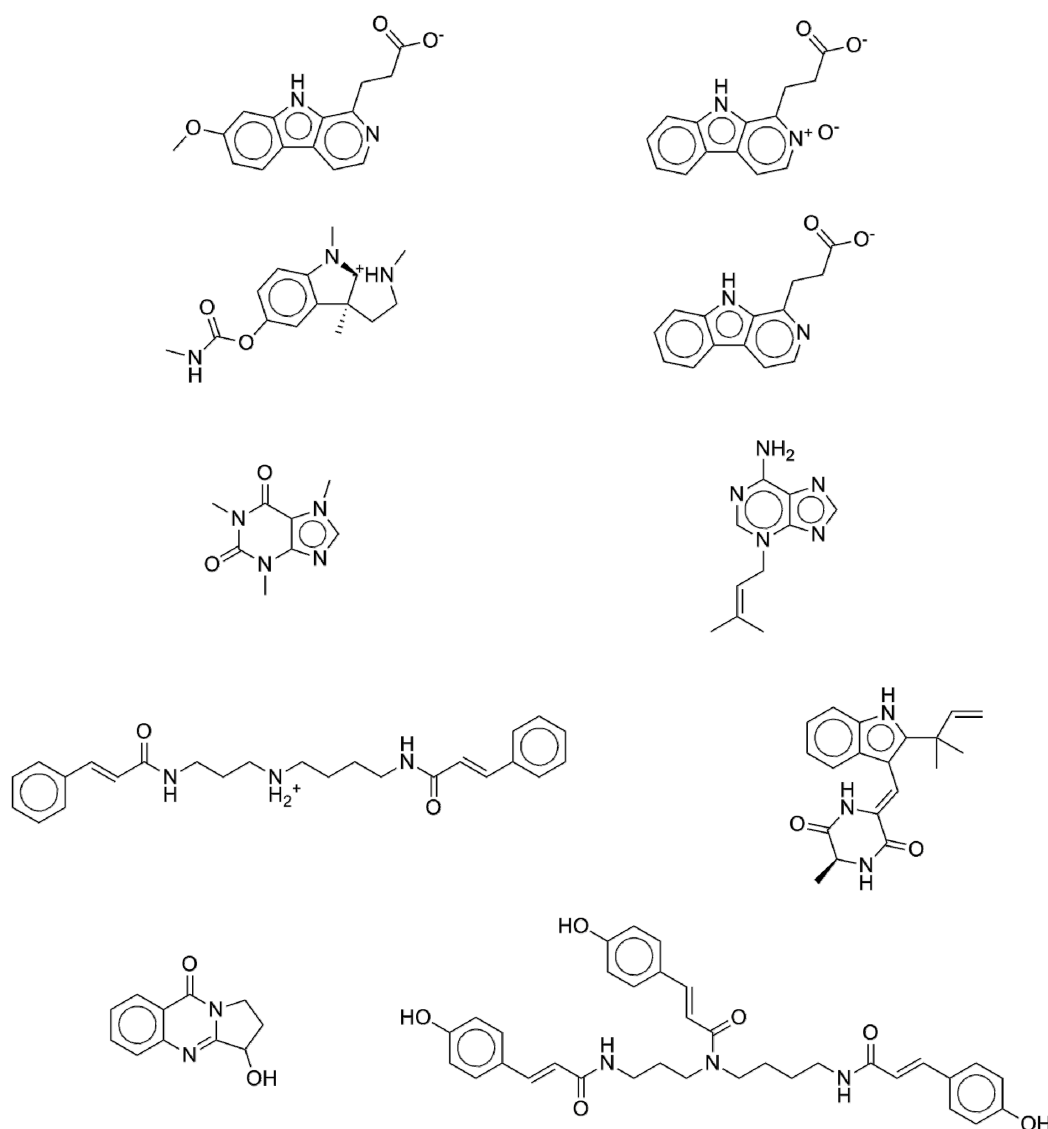

Figure 9. Some of the natural products that scored highly as potential anti-diabetic drug candidates according to our ISE-based, anti-diabetic activity indexing model.

3. Materials and Methods

For modeling purposes, we used a set of 97 anti-diabetic drugs (presented in SMILES format in the Supporting Information (Table S1)) to represent the active domain. Diversity within the anti-diabetic drugs is shown in Figure 1A. Furthermore, a set composed of 2892 natural products, sharing intermolecular similarity values <0.9, was used to represent the inactive domain. This is justified, since the prediction's model used for virtual screening should cover the properties of chemicals from the screened database, and this database, which was prepared by collecting phytochemicals isolated from more than 800 different plants, is distributed worldwide and is available for purchase from AnalytiCon Discovery company that is located in Postdam, Germany (www.ac-discovery.com). The diversity among the natural products is shown in Figure 1B. It should be noted that the 2892 natural products are putatively inactive and probably contain few active chemical. The usage of a putative inactive set of chemicals that might be 'contaminated' with a few active chemicals is discussed in our drug-likeness paper [46]. The ACD database was used as a representative set of non-drugs, although it contains about 1% drugs. The filter-based prediction technique was found to be less sensitive to the 'contamination' of the inactive set by 1–2% active compounds.

The physico-chemical properties of all of the chemicals in both databases were calculated using Molecular Operating Environment (MOE) software, Version 2009.10 (http://www.chemcomp.com). The 2D descriptors were based on calculated properties such as molecular weight, log P, H-bond donors/acceptors, solubility, total charge and charge distribution, the types and number of atoms, and so forth [65]. To assess and validate the predictability of the model, the datasets of active/inactive ligands were split into 66.7% for the training set and 33.3% for the test set; an in-house random picking module generated these sets.

The iterative stochastic elimination (ISE) algorithm [46] was implemented to construct models tailored to indexing natural products for potential anti-diabetic activity. Optimal differentiation between active and inactive chemicals was attained by searching, in multivariable space, for the best sets of descriptors ('variables') and the best ranges for all descriptors in each set, capable of distinguishing between active and inactive chemicals. Since chemical descriptors generally interact with each other, changes in the range of one descriptor could have an effect on the best range of another descriptor, and the optimization process is obliged to take into consideration all descriptors in the set at one time to attain the best set of filters. The flowchart of the modeling process is shown in Figure 2. For further details on applying the ISE algorithm to obtain the best ranges for a set of descriptors, and for the optimization process, see our previously-reported studies [46,55].

Employing a 'combined filters approach' increases the discrimination power and makes it possible to attach to each chemical a molecular bioactivity index (MBI) that correlates with the chance of a molecule's being bioactive. The MBI concept is based on the assumption that a bioactive ligand would pass more 'filters', while an inactive molecule would pass a minimal number of filters. This is the basis for the construction of the MBI index, which is composed of the contribution of the number of filters passed by a molecule to that molecule's overall quality of potential bioactivity.

$$MBI = \frac{\sum_{i=1}^{n} \left(\delta Ai \frac{PAi}{PNAi} - \delta NAi \frac{NNAi}{NAi} \right)}{n} \tag{1}$$

The value of the delta function δAi is 0 (zero) if the molecule is inactive according to the currently calculated filter i and 1 if it is bioactive according to that filter. Similarly, the value of the delta function δ_{NAi} is 1 if it is inactive according to filter i, and 0 if it is bioactive according to that filter. PAi is the percentage of bioactive molecules that are predicted to be 'bioactives' according to filter i ('true positives'), while $PNAi$ is the percentage of false positives, i.e., inactive molecules that are predicted to be bioactives according to filter i. NAi is the percent of bioactives identified to be inactives according to the current filter ('false negatives'), and N_{NAi} is the percent of inactives identified as such by the current filter (i.e., 'true negatives'). The quotient $PAi/PNAi$ may be regarded as an 'efficiency factor' of filter i for bioactives, while the quotient $NAi/NNAi$ is an 'efficiency factor' for misidentifying inactives.

Various parameters such as the enrichment factor, Matthews correlation coefficient (MCC), the ROC curve and the area under ROC curve (AUC) were used to assess the quality of the prediction model. In the ROC curve, sensitivity (the true positive rate) is plotted as a function of the false positive rate. The AUC (area under the ROC curve) is a measure of how well the prediction model can distinguish between two chemicals, one active and one inactive.

4. Conclusions

Since diabetes mellitus (DM) is still a leading disease with lethal concomitant complications affecting people worldwide and is associated with an unmet need for developing novel anti-diabetic drugs, several research groups in academia and industry are using in silico techniques to facilitate the discovery of novel anti-diabetic drug candidates, while aiming to save time and costs. We have constructed a prediction model using a set of 97 approved anti-diabetic drugs to represent the active domain and a set of 2892 natural products to represent the inactive domain. It is worth noting that only a few out of the 2892 natural products could have been expected to be anti-diabetic, but the effect of that assumption on the quality of the prediction model is assumed to be negligible. To obtain accurate predictive models for virtual screening purposes, the modeling process should use sets of chemicals that cover the space of the properties of the objects in the screened database. Consequently, we had to select, as the inactive set, objects with the same 'property space' as the screened objects. The optimization technique used in this study to index natural products for their potential anti-diabetic bioactivity was the iterative stochastic elimination algorithm. A highly discriminative and robust model was obtained with the area under the curve (AUC) >0.96, indicating a very good prediction model. Upon application of the proposed anti-diabetic activity indexing model to the virtual screening of a set of chemicals with a mix ratio of 1:1000 active-to-inactive compounds, 65% of the anti-diabetic drugs were captured in the top one percent of the screened compounds, compared to 1% in the random model. Some of the natural products that got a high score as anti-diabetic drug candidates are disclosed and presented in Figure 9. A search of the literature revealed that one of the high-scoring phytochemicals (caffeine) has already been tested and reported as an active anti-diabetic molecule, capable to decrease blood glucose levels. The other nine phytochemicals await evaluation in the wet lab for their anti-diabetic activity.

Author Contributions: All five authors contributed extensively to the work presented in this paper. A.R. conceived of the study and edited the manuscript. M.R. created the indexing model and performed the analysis under the supervision of A.R. M.Z., N.Z and M.F. interpreted the data and wrote the manuscript.

Conflicts of Interest: The authors declare that they have no conflict of interests.

References

1. Alberti, K.G.; Zimmet, P.Z. Definition, diagnosis and classification of diabetes mellitus and its complications. Part 1: Diagnosis and classification of diabetes mellitus provisional report of a WHO consultation. *Diabet. Med.* **1998**, *15*, 539–553. [CrossRef]
2. Whiting, D.R.; Guariguata, L.; Weil, C.; Shaw, J. IDF diabetes atlas: Global estimates of the prevalence of diabetes for 2011 and 2030. *Diabetes Res. Clin. Pract.* **2011**, *94*, 311–321. [CrossRef] [PubMed]
3. International Diabetes Federation. *IDF Diabetes Atlas*, 7th ed.; International Diabetes Federation: Brussels, Belgium, 2015.
4. Kahn, S.E.; Cooper, M.E.; Del Prato, S. Pathophysiology and treatment of type 2 diabetes: Perspectives on the past, present, and future. *Lancet* **2014**, *383*, 1068–1083. [CrossRef]

5. World Health Organization. Global Health Estimates (GHE), 2000–2015 estimates. Available online: http://www.who.int/mediacentre/factsheets/fs312/en/ (accessed on 12 August 2017).

6. Muoio, D.M.; Newgard, C.B. Mechanisms of disease:Molecular and metabolic mechanisms of insulin resistance and beta-cell failure in type 2 diabetes. *Nat. Rev. Mol. Cell Biol.* **2008**, *9*, 193–205. [CrossRef] [PubMed]

7. Fonseca, V.A. Defining and characterizing the progression of type 2 diabetes. *Diabetes Care* **2009**, *32* (Suppl. 2), S151–S156. [CrossRef] [PubMed]

8. Pessin, J.E.; Saltiel, A.R. Signaling pathways in insulin action: Molecular targets of insulin resistance. *J. Clin. Investig.* **2000**, *106*, 165–169. [CrossRef] [PubMed]

9. DeFronzo, R.A. Lilly lecture 1987. The triumvirate: Beta-cell, muscle, liver. A collusion responsible for NIDDM. *Diabetes* **1988**, *37*, 667–687. [CrossRef] [PubMed]

10. Defronzo, R.A. Banting Lecture. From the triumvirate to the ominous octet: A new paradigm for the treatment of type 2 diabetes mellitus. *Diabetes* **2009**, *58*, 773–795. [CrossRef] [PubMed]

11. Bavenholm, P.N.; Kuhl, J.; Pigon, J.; Saha, A.K.; Ruderman, N.B.; Efendic, S. Insulin resistance in type 2 diabetes: Association with truncal obesity, impaired fitness, and atypical malonyl coenzyme A regulation. *J. Clin. Endocrinol. Metab.* **2003**, *88*, 82–87. [CrossRef] [PubMed]

12. Gustat, J.; Srinivasan, S.R.; Elkasabany, A.; Berenson, G.S. Relation of self-rated measures of physical activity to multiple risk factors of insulin resistance syndrome in young adults: The Bogalusa Heart Study. *J. Clin. Epidemiol.* **2002**, *55*, 997–1006. [CrossRef]

13. Kahn, R.; Buse, J.; Ferrannini, E.; Stern, M.; American Diabetes Association; European Association for the Study of Diabetes. The metabolic syndrome: Time for a critical appraisal: Joint statement from the American Diabetes Association and the European Association for the Study of Diabetes. *Diabetes Care* **2005**, *28*, 2289–2304. [CrossRef] [PubMed]

14. Wingard, D.L.; Barrett-Connor, E.; Criqui, M.H.; Suarez, L. Clustering of heart disease risk factors in diabetic compared to nondiabetic adults. *Am. J. Epidemiol.* **1983**, *117*, 19–26. [CrossRef] [PubMed]

15. Miranda, P.J.; DeFronzo, R.A.; Califf, R.M.; Guyton, J.R. Metabolic syndrome: Evaluation of pathological and therapeutic outcomes. *Am. Heart J.* **2005**, *149*, 20–32. [CrossRef] [PubMed]

16. Miranda, P.J.; DeFronzo, R.A.; Califf, R.M.; Guyton, J.R. Metabolic syndrome: Definition, pathophysiology, and mechanisms. *Am. Heart J.* **2005**, *149*, 33–45. [CrossRef] [PubMed]

17. Adler, A.I.; Stratton, I.M.; Neil, H.A.; Yudkin, J.S.; Matthews, D.R.; Cull, C.A.; Wright, A.D.; Turner, R.C.; Holman, R.R. Association of systolic blood pressure with macrovascular and microvascular complications of type 2 diabetes (UKPDS 36): Prospective observational study. *BMJ* **2000**, *321*, 412–419. [CrossRef] [PubMed]

18. Reaven, G.M. Role of insulin resistance in human disease (syndrome X): An expanded definition. *Annu. Rev. Med.* **1993**, *44*, 121–131. [CrossRef] [PubMed]

19. Inzucchi, S.E.; Bergenstal, R.M.; Buse, J.B.; Diamant, M.; Ferrannini, E.; Nauck, M.; Peters, A.L.; Tsapas, A.; Wender, R.; Matthews, D.R. Management of hyperglycemia in type 2 diabetes, 2015: A patient-centered approach: Update to a position statement of the American Diabetes Association and the European Association for the Study of Diabetes. *Diabetes Care* **2015**, *38*, 140–149. [CrossRef] [PubMed]

20. Smith, R.J.; Nathan, D.M.; Arslanian, S.A.; Groop, L.; Rizza, R.A.; Rotter, J.I. Individualizing therapies in type 2 diabetes mellitus based on patient characteristics: What we know and what we need to know. *J. Clin. Endocrinol. Metab.* **2010**, *95*, 1566–1574. [CrossRef] [PubMed]

21. Stratton, I.M.; Adler, A.I.; Neil, H.A.; Matthews, D.R.; Manley, S.E.; Cull, C.A.; Hadden, D.; Turner, R.C.; Holman, R.R. Association of glycaemia with macrovascular and microvascular complications of type 2 diabetes (UKPDS 35): Prospective observational study. *BMJ* **2000**, *321*, 405–412. [CrossRef] [PubMed]

22. UK Prospective Diabetes Study (UKPDS) Group. Intensive blood-glucose control with sulphonylureas or insulin compared with conventional treatment and risk of complications in patients with type 2 diabetes (UKPDS 33). *Lancet* **1998**, *352*, 837–853.

23. Holman, R.R.; Paul, S.K.; Bethel, M.A.; Matthews, D.R.; Neil, H.A. 10-year follow-up of intensive glucose control in type 2 diabetes. *New Engl. J. Med.* **2008**, *359*, 1577–1589. [CrossRef] [PubMed]

24. Hu, F.B.; Manson, J.E.; Stampfer, M.J.; Colditz, G.; Liu, S.; Solomon, C.G.; Willett, W.C. Diet, lifestyle, and the risk of type 2 diabetes mellitus in women. *New Engl. J. Med.* **2001**, *345*, 790–797. [CrossRef] [PubMed]

25. Eriksson, A.K.; van den Donk, M.; Hilding, A.; Ostenson, C.G. Work stress, sense of coherence, and risk of type 2 diabetes in a prospective study of middle-aged Swedish men and women. *Diabetes Care* **2013**, *36*, 2683–2689. [CrossRef] [PubMed]

26. Ostenson, C.G.; Hilding, A.; Grill, V.; Efendic, S. High consumption of smokeless tobacco ('snus') predicts increased risk of type 2 diabetes in a 10-year prospective study of middle-aged Swedish men. *Scand. J. Public Health* **2012**, *40*, 730–737. [CrossRef] [PubMed]

27. Helmrich, S.P.; Ragland, D.R.; Leung, R.W.; Paffenbarger, R.S., Jr. Physical activity and reduced occurrence of non-insulin-dependent diabetes mellitus. *New Engl. J. Med.* **1991**, *325*, 147–152. [CrossRef] [PubMed]

28. Wikner, C.; Gigante, B.; Hellenius, M.L.; de Faire, U.; Leander, K. The risk of type 2 diabetes in men is synergistically affected by parental history of diabetes and overweight. *PLoS ONE* **2013**, *8*, e61763. [CrossRef] [PubMed]

29. Hilding, A.; Eriksson, A.K.; Agardh, E.E.; Grill, V.; Ahlbom, A.; Efendic, S.; Ostenson, C.G. The impact of family history of diabetes and lifestyle factors on abnormal glucose regulation in middle-aged Swedish men and women. *Diabetologia* **2006**, *49*, 2589–2598. [CrossRef] [PubMed]

30. Prasad, R.B.; Groop, L. Genetics of type 2 diabetes-pitfalls and possibilities. *Genes* **2015**, *6*, 87–123. [CrossRef] [PubMed]

31. Bailey, C.J.; Day, C. Traditional plant medicines as treatments for diabetes. *Diabetes Care* **1989**, *12*, 553–564. [CrossRef] [PubMed]

32. Clarence, O.K.; Donatus, B.C.; Samuel, E.C. Prophylaxis and treatment of types 1 and 2 diabetes mellitus. *Int. J. Dis. Disord.* **2014**, *2*, 65–73.

33. Ivorra, M.D.; Paya, M.; Villar, A. A review of natural products and plants as potential antidiabetic drugs. *J. Ethnopharmacol.* **1989**, *27*, 243–275. [CrossRef]

34. Rai, M.K. A review on some antidiabetic plants of India. *Anc. Sci. Life* **1995**, *14*, 168–180. [PubMed]

35. Zatsepin, M.; Mattes, A.; Rupp, S.; Finkelmeier, D.; Basu, A.; Burger-Kentischer, A.; Goldblum, A. Computational Discovery and Experimental Confirmation of TLR9 Receptor Antagonist Leads. *J. Chem. Inf. Model.* **2016**, *56*, 1835–1846. [CrossRef] [PubMed]

36. Basu, A.; Sohn, Y.S.; Alyan, M.; Nechushtai, R.; Domb, A.J.; Goldblum, A. Discovering Novel and Diverse Iron-Chelators in Silico. *J. Chem. Inf. Model.* **2016**, *56*, 2476–2485. [CrossRef] [PubMed]

37. Pappalardo, M.; Shachaf, N.; Basile, L.; Milardi, D.; Zeidan, M.; Raiyn, J.; Guccione, S.; Rayan, A. Sequential application of ligand and structure based modeling approaches to index chemicals for their hH4R antagonism. *PLoS ONE* **2014**, *9*, e109340. [CrossRef] [PubMed]

38. Zaid, H.; Raiyn, J.; Osman, M.; Falah, M.; Srouji, S.; Rayan, A. In silico modeling techniques for predicting the tertiary structure of human H4 receptor. *Front Biosci.* **2016**, *21*, 597–619.

39. Schuller, A.; Schneider, G. Identification of hits and lead structure candidates with limited resources by adaptive optimization. *J. Chem. Inf. Model.* **2008**, *48*, 1473–1491. [CrossRef] [PubMed]

40. Hao, M.; Zhang, S.; Qiu, J. Toward the Prediction of FBPase Inhibitory Activity Using Chemoinformatic Methods. *Int. J. Mol. Sci.* **2012**, *13*, 7015–7037. [CrossRef] [PubMed]

41. Lusci, A.; Pollastri, G.; Baldi, P. Deep architectures and deep learning in chemoinformatics: The prediction of aqueous solubility for drug-like molecules. *J. Chem. Inf. Model.* **2013**, *53*, 1563–1575. [CrossRef] [PubMed]

42. Heikamp, K.; Bajorath, J. Comparison of confirmed inactive and randomly selected compounds as negative training examples in support vector machine-based virtual screening. *J. Chem. Inf. Model.* **2013**, *53*, 1595–1601. [CrossRef] [PubMed]

43. Shen, M.; Beguin, C.; Golbraikh, A.; Stables, J.P.; Kohn, H.; Tropsha, A. Application of predictive QSAR models to database mining: Identification and experimental validation of novel anticonvulsant compounds. *J. Med. Chem.* **2004**, *47*, 2356–2364. [CrossRef] [PubMed]

44. Rayan, A.; Falah, M.; Mawasi, H.; Raiyn, N. Assessing drugs for their cardio-toxicity. *Lett. Drug Des. Discov.* **2010**, *7*, 409–414. [CrossRef]

45. Deeb, O.; Jawabreh, S.; Goodarzi, M. Exploring QSARs of vascular endothelial growth factor receptor-2 (VEGFR-2) tyrosine kinase inhibitors by MLR, PLS and PC-ANN. *Curr. Pharm. Des.* **2013**, *19*, 2237–2244. [CrossRef] [PubMed]

46. Rayan, A.; Marcus, D.; Goldblum, A. Predicting oral druglikeness by iterative stochastic elimination. *J. Chem. Inf. Model.* **2010**, *50*, 437–445. [CrossRef] [PubMed]

47. Mussa, H.Y.; Hawizy, L.; Nigsch, F.; Glen, R.C. Classifying large chemical data sets: Using a regularized potential function method. *J. Chem. Inf. Model.* **2011**, *51*, 4–14. [CrossRef] [PubMed]

48. Pappalardo, M.; Rayan, M.; Abu-Lafi, S.; Leonardi, M.E.; Milardi, D.; Guccione, S.; Rayan, A. Homology-based Modeling of Rhodopsin-like Family Members in the Inactive State: Structural Analysis and Deduction of Tips for Modeling and Optimization. *Mol. Inform.* **2017**, *36*. [CrossRef] [PubMed]

49. Shahaf, N.; Pappalardo, M.; Basile, L.; Guccione, S.; Rayan, A. How to Choose the Suitable Template for Homology Modelling of GPCRs: 5-HT7 Receptor as a Test Case. *Mol. Inform.* **2016**, *35*, 414–423. [CrossRef] [PubMed]

50. Garcia-Sosa, A.T.; Oja, M.; Hetenyi, C.; Maran, U. DrugLogit: Logistic discrimination between drugs and nondrugs including disease-specificity by assigning probabilities based on molecular properties. *J. Chem. Inf. Model.* **2012**, *52*, 2165–2180. [CrossRef] [PubMed]

51. Garcia-Sosa, A.T.; Oja, M.; Hetenyi, C.; Maran, U. Disease-Specific Differentiation Between Drugs and Non-Drugs Using Principal Component Analysis of Their Molecular Descriptor Space. *Mol. Inform.* **2012**, *31*, 369–383. [CrossRef] [PubMed]

52. Garcia-Sosa, A.T.; Maran, U.; Hetenyi, C. Molecular property filters describing pharmacokinetics and drug binding. *Curr. Med. Chem.* **2012**, *19*, 1646–1662. [CrossRef] [PubMed]

53. Rayan, A.; Noy, E.; Chema, D.; Levitzki, A.; Goldblum, A. Stochastic algorithm for kinase homology model construction. *Curr. Med. Chem.* **2004**, *11*, 675–692. [CrossRef] [PubMed]

54. Rayan, A. The utility of Intelligent Learning Engine in Drug Discovery Informatics. *Proc. Br. Pharmacol. Soc.* **2010**, *7*, 26.

55. Rayan, A.; Falah, M.; Raiyn, J.; Da'adoosh, B.; Kadan, S.; Zaid, H.; Goldblum, A. Indexing molecules for their hERG liability. *Eur. J. Med. Chem.* **2013**, *65*, 304–314. [CrossRef] [PubMed]

56. Glick, M.; Goldblum, A. A novel energy-based stochastic method for positioning polar protons in protein structures from X-rays. *Proteins* **2000**, *38*, 273–287. [CrossRef]

57. Glick, M.; Rayan, A.; Goldblum, A. A stochastic algorithm for global optimization and for best populations: A test case of side chains in proteins. *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 703–708. [CrossRef] [PubMed]

58. Michaeli, A.; Rayan, A. Modeling Ensembles of Loop Conformations by Iterative Stochastic Elimination. *Lett. Drug Des. Discov.* **2016**, *13*, 1–6. [CrossRef]

59. Rayan, A.; Senderowitz, H.; Goldblum, A. Exploring the conformational space of cyclic peptides by a stochastic search method. *J. Mol. Graph. Model.* **2004**, *22*, 319–333. [CrossRef] [PubMed]

60. Zaid, H.; Raiyn, J.; Nasser, A.; Saad, B.; Rayan, A. Physicochemical properties of natural based products versus synthetic chemicals. *Open Nutraceuticals J.* **2010**, *3*, 194–202. [CrossRef]

61. Frank, A.; Abu-Lafi, S.; Adawi, A.; Schwed, J.S.; Stark, H.; Rayan, A. From medicinal plant extracts to defined chemical compounds targeting the histamine H4 receptor: Curcuma longa in the treatment of inflammation. *Inflamm. Res.* **2017**, 1–7. [CrossRef] [PubMed]

62. Kacergius, T.; Abu-Lafi, S.; Kirkliauskiene, A.; Gabe, V.; Adawi, A.; Rayan, M.; Qutob, M.; Stukas, R.; Utkus, A.; Zeidan, M.; et al. Inhibitory capacity of Rhus coriaria L. extract and its major component methyl gallate on Streptococcus mutans biofilm formation by optical profilometry: Potential applications for oral health. *Mol. Med. Rep.* **2017**, *16*, 949–956. [CrossRef] [PubMed]

63. Lipinski, C.A.; Lombardo, F.; Dominy, B.W.; Feeney, P.J. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv. Drug Deliv. Rev.* **2001**, *46*, 3–26. [CrossRef]

64. Hann, M.M.; Oprea, T.I. Pursuing the leadlikeness concept in pharmaceutical research. *Curr. Opin. Chem. Biol.* **2004**, *8*, 255–263. [CrossRef] [PubMed]

65. Chemical Computing Group. QuaSAR-Descriptor. Available online: http://www.chemcomp.com/journal/descr.htm (accessed on 12 August 2017).

66. Fang, C.Y.; Wang, X.J.; Huang, Y.W.; Hao, S.M.; Sheng, J. Caffeine is responsible for the bloodglucose-lowering effects of green tea and Puer tea extractsin BALB/c mice. *Chin. J. Nat. Med.* **2015**, *13*, 595–601. [CrossRef]

67. Matsuda, Y.; Kobayashi, M.; Yamauchi, R.; Ojika, M.; Hiramitsu, M.; Inoue, T.; Katagiri, T.; Murai, A.; Horio, F. Coffee and caffeine improve insulin sensitivity and glucose tolerance in C57BL/6J mice fed a high-fat diet. *Biosci. Biotechnol. Biochem.* **2011**, *75*, 2309–2315. [CrossRef] [PubMed]
68. Ozmen, O.; Topsakal, S.; Haligur, M.; Aydogan, A.; Dincoglu, D. Effects of Caffeine and Lycopene in Experimentally Induced Diabetes Mellitus. *Pancreas* **2016**, *45*, 579–583. [CrossRef] [PubMed]

Sample Availability: The natural products database is available from the corresponding author upon request.