RSC Advances

PAPER

View Article Online
View Journal | View Issue

Cite this: *RSC Adv.*, 2021, **11**, 5172

DiaNat-DB: a molecular database of antidiabetic compounds from medicinal plants†

Abraham Madariaga-Mazón, *[a] José J. Naveja,[a] José L. Medina-Franco,[b] Karla O. Noriega-Colima[a] and Karina Martinez-Mayorga *[a]

Natural products are an invaluable source of molecules with a large variety of biological activities. Interest in natural products in drug discovery is documented in an increasing number of publications of bioactive secondary metabolites. Among those, medicinal plants are one of the most studied for this endeavor. An ever thriving area of opportunity within the field concerns the discovery of antidiabetic natural products. As a result, a vast amount of secondary metabolites are isolated from medicinal plants used against diabetes mellitus but whose information has not been organized systematically yet. Several research articles enumerate antidiabetic compounds, but the lack of a chemical database for antidiabetic metabolites limits their application in drug development. In this work, we present DiaNat-DB, a comprehensive collection of 336 molecules from medicinal plants reported to have *in vitro* or *in vivo* antidiabetic activity. We also discuss a chemoinformatic analysis of DiaNat-DB to compare antidiabetic drugs and natural product databases. To further explore the antidiabetic chemical space based on DiaNat compounds, we searched for analogs in ZINC15, an extensive database listing commercially available compounds. This work will help future analyses, design, and development of new antidiabetic drugs. DiaNat-DB and its ZINC15 analogs are freely available at http://rdu.iquimica.unam.mx/handle/20.500.12214/1186.

Received 12th December 2020
Accepted 15th January 2021

DOI: 10.1039/d0ra10453a

rsc.li/rsc-advances

1. Introduction

Diabetes mellitus (DM) is a chronic metabolic disease characterized by a hyperglycemic physiological state. Insulin deficiency or insulin resistance leads to hyperglycemia.[1] The International Diabetes Federation (IDF) estimates that 463 million people aged 20–79 have diabetes mellitus globally. The projection made in 2010 for 2025 was 438 million. Thus, we are already 25 million over the prediction, with five more years to go. The IDF estimates that by 2030 there will be 578 million people with diabetes worldwide.[2,3] There are two main diabetes mellitus variants; type 1 (T1DM) and type 2 (T2DM). T1DM is caused by autoimmune destruction of pancreatic beta-cells, producing endogen insulin depletion. T2DM is the most common form of DM; its driving physiopathological mechanism is insulin resistance with an inadequate compensatory insulin secretion.[1] Chronic hyperglycemia causes several life-threatening complications, such as retinopathy, nephropathy, and neuropathy; major complications include accelerated cardiovascular disease leading to myocardial infarction and cerebrovascular disease.[4] Nonetheless, prompt glucose, blood pressure, and cholesterol blood levels control ameliorates most DM complications.[4,5] The pharmacological treatment in T2DM focuses on controlling fasting and postprandial plasma glucose levels, as well as counteracting the three main metabolic alterations: decreased pancreatic β-cell function, elevated hepatic glucose production, and insulin resistance.[6] Most drugs used to treat T2DM produce many adverse effects, such as mild to severe hypoglycemia, weight gain, diarrhea, nausea, lactic acidosis, and even heart failure. Although antidiabetic drugs offer a wide range of metabolic actions to reduce hyperglycemia and its long-term complications,[7] the life expectancy of diabetic patients is still significantly reduced.[8] Moreover, T2DM treatment must be individualized and thus requires a variety of available treatments.[7] Therefore, the discovery of new compounds with suitable pharmacological profiles are still needed.

The World Health Organization (WHO) defines a medicinal plant as any plant species containing substances that have therapeutic usefulness or whose secondary metabolites can serve as precursors for the synthesis of new drugs.[9] Traditional medicine is a mainstream therapeutic resource of healthcare in developing countries, and the use of complementary and

[a]*Instituto de Química, Universidad Nacional Autónoma de México, Mexico City 04510, Mexico. E-mail: amadariaga@iquimica.unam.mx; kmtzm@unam.mx; Tel: +52 55 56224770 ext. 46614*

[b]*DIFACQUIM Research Group, Department of Pharmacy, School of Chemistry, Universidad Nacional Autónoma de México, Mexico City 04510, Mexico*

† Electronic supplementary information (ESI) available: DiaNat-DB, ZINC analogs, and R-group tables for cores in ZINC are available for download at http://rdu.iquimica.unam.mx/handle/20.500.12214/1186. See DOI: 10.1039/d0ra10453a

Table 1 Available information on natural products with antidiabetic activity

Web server/database	Description	Comments	Reference
NeMedPlant	It contains information about medicinal plants from northwestern India and their active components.	The information provided includes medicinal plants used to treat other diseases than diabetes (tuberculosis, cancer, *etc.*). The chemical structure of bioactive compounds is not provided. Freely searchable database.	16
DIA-DB	Web server for the prediction of bioactive compounds for the treatment of diabetes using comparison by similarity or inverse virtual screening.	The database available provides only information on the drugs approved for the treatment of DM but does not contain information on the active compounds of natural products. Freely searchable database.	17
ADNCD	It contains information on natural antidiabetic compounds classified on the basis of their mode of action.	For some entries, the database provides information on physicochemical properties and risks of toxicity. Freely searchable database.	18
Database on antidiabetic indigenous plants of Tamil Nadu, India	It contains information about medicinal plants used in the treatment of diabetes mellitus and its complications.	In some instances, there is no information on the bioactive compounds, or the chemical structure cannot be found. Descriptive database, not available for search or download.	19

alternative medicine is hastily increasing in developed countries.[10] In fact, natural products represent primary starting points in drug discovery campaigns.[11] Medicinal plants have played an essential role in the treatment of chronic diseases, including T2DM.[12] For instance, several secondary metabolites have hypoglycemic activity, including phenolic compounds (phenolic acids, xanthones, and flavonoids) and alkaloids.[13–15] Nonetheless, the available compendia on antidiabetic natural products are limited to review articles and databases focused on medicinal plants, with little information on the chemical structures of active metabolites (Table 1). To fill this gap, we constructed and curated DiaNat-DB, the first publicly available database with integrated molecular information on antidiabetic natural products.

2. Materials and methods

2.1. Literature search

We conducted a literature search using the scientific search engines SciFinder, Scopus, ScienceDirect, Pubmed, and the *Atlas de la Medicina Tradicional*. The revision included documents published until 2019, and updates will be performed on a biannual basis. The keywords employed were "diabetes", "hypoglycemic agents", "hyperglycemia", "medicinal plants", and "natural products". The search was limited to original articles and reviews of bioactive compounds from medicinal plants. Articles without explicit information on isolated bioactive compounds were excluded. Our search criteria was focused on compounds with antidiabetic activity from plants used in folk medicine, regardless of the biological target. Thus, the search was not restricted to or focused on any biological target. Every single compound included in the database has evidence of antidiabetic activity in human consumption as well as antidiabetic activity in animal models (*in vivo*) or *in vitro* assays. Moreover, the *in vivo*/*in vitro* activity is reported for the isolated compound which is chemically characterized. If a compound did not fulfill these criteria it was not included in the database.

2.2. Database compilation and curation

The chemical structures were prepared with ChemDraw Direct (2.0.01262) and converted to SMILES representation (Simplified Molecular Input Line Entry Specification).[20] Data curation was performed in MOE (Molecular Operating Environment v.2019). It consisted of assigning protonation states at physiological pH (7.4), adding explicit hydrogen atoms; conserving the structures' chirality as defined in the source publication, and rebuilding structures that were not correctly constructed. We also conducted the energy-minimization of the structures using the MMFF94 force field.

2.3. Reference databases

The diversity of DiaNat-DB was compared to reference databases from DrugBank 5.0:[21] drugs approved against T2DM (DM-Ref), and all FDA-approved small molecules (FDA).

2.4. Descriptors, data analysis and visualization

We computed six physicochemical properties of pharmaceutical relevance for the bioactive compounds. The properties are

associated with the empirical drug-likeness rules of Lipinski and Veber *et al.*[22,23] and are frequently used to profile compound databases for drug discovery applications:

- molecular weight (Da)
- partition coefficient (log P)
- number of hydrogen acceptor atoms
- number of hydrogen donors atoms
- number of rotatable bonds
- sp^3 atoms (although not considered by the rule of five, it is useful to measure molecular complexity[24]), and it has also been proposed as a measure of drug-likeness.[25]

The six properties were calculated with DataWarrior.

2.5. Scaffold content, diversity and structure-similarity analyses

We conducted a Bemis–Murcko scaffold content and similarity analysis for FDA and DiaNat-DB. Briefly, the Bemis–Murcko scaffold of a chemical structure is the main core structure that remains after removing the side chains.[26] This allowed similarity-searching of scaffolds between both databases (ESI†). Structure-similarity was measured with the Tanimoto coefficient of ECFP-4.

2.6. Analog enrichment

Natural products are commonly challenging to synthesize, and the yields from the extracts are typically low. An alternative way to explore the bioactivity of more accessible compounds is through the search for structural analogs. A recent methodology for this purpose is the exploration of constellation plots. Constellation plots are graphical representations of the analogs series in the chemical space; their generation requires chemical fingerprints (Morgan FP in this case) and a dimensionality reduction method (t-SNE[27]). The reader interested in more details on analog series analysis is referred to citations.[28–31]

We followed the methodology described in ref. [28] and [29] to find Retrosynthetic Combinatorial Analysis Procedure (RECAP) chemical analogs in ZINC15 (ref. [32]) for molecules in DiaNat-DB. We also constructed constellation plots[31] and R tables[30] using in-house scripts. For methodological details we refer the interested reader to the corresponding citations. A "core" is defined as a substructure of a molecule obtained by RECAP retrosynthesis rules and retains at least two-thirds of the total heavy atoms. The definition allows more than one core per molecule. Analog series are defined by cores sharing molecules. Of note, stereochemistry was disregarded for core analysis.

3. Results

3.1. Data collection and selection

Our systematic search initially yielded approximately 9900 peer-reviewed publications. After analyzing the abstracts and applying the exclusion criteria described in the Methods section, 959 articles remained. Full-text revision showed that about 200 documents contained data from medicinal plants used for treating T2DM.

DiaNat-DB contains phenolic compounds, alkaloids, and terpenes (see Table 2), three of the main classes of secondary metabolites.

We broadly classified the metabolites as "hypoglycemic" or "antihyperglycemic" based on the reported activity. In some cases, the compound's information did not contain any detail about the type of activity, so it was classified generically as "antidiabetic". Additionally, a set of compounds was classified as "focused on complications" since it targets the complications derived from DM.

3.2. Database generation

DiaNat-DB contains 336 unique compounds, annotated with the following information: molecular structure (in SMILES), PubChemID, compound identifier (ID), compounds' name, source plant, medicinal use of the plant, genus and family, the reported activity of the compound, details of the mechanism of action (when available), the plant geographical information, and bibliographic citation. The dataset is provided in the ESI.† DiaNat-DB contributes to the natural product databases available in the public domain.[33,34]

Table 2 General classes of compounds contained in DiaNat-DB, examples, and natural origin of isolation

General class			Exemplary compounds	Natural origin
Phenolic compounds	Phenolics acids		Caffeic acid, gallic acid, chlorogenic acid	*Xanthium strumarium*, *Malus domestica*, *Passiflora ligularis*
	Xanthones		Mangiferin	*Mangifera indica*, *Salacia chinensis*
	Flavonoids	Flavones	Acacetin, diosmetin, luteolin	*Anoda cristata*, *Chamaemelum nobile*
		Flavanols	Catechin, epigallocatechin gallate (EGCG), (−)-epicatechin	*Camellia sinensis*, *Astrocaryum aculeatum*
		Flavonols	Quercetin, rutin, kaempferol, kaempferitrin, myricetin	*Dillenia indica*, *Morus alba*, *Pterogyne nitens*, *Abelmoschus moschatus*
		Flavanones	Naringenin, hesperidin	*Citrus sinensis*, *Carissa carandas*
		Isoflavones	Genistein, daidzein	*Glycine max*, *Pueraria lobata*
		Anthocyanidins	Cyanidin, cyanidin-3-glucoside, pelargonidin	*Aristotelia chilensis*, *Vaccinium oxycoccos*, *Ribes nigrum*, *Anagallis monelli*
Alkaloids			Ajmaline, berberine, palmatine, trigonelline	*Rauvolfia serpentina*, *Coptis chinensis*, *Bauhinia candicans*
Terpenes		Monoterpenes	Linalool, limonene	*Pelargonium graveolens*, *Aegle marmelos*
		Sesquiterpenes	Costunolide, ar-turmerone	*Costus speciosus*, *Curcuma longa*

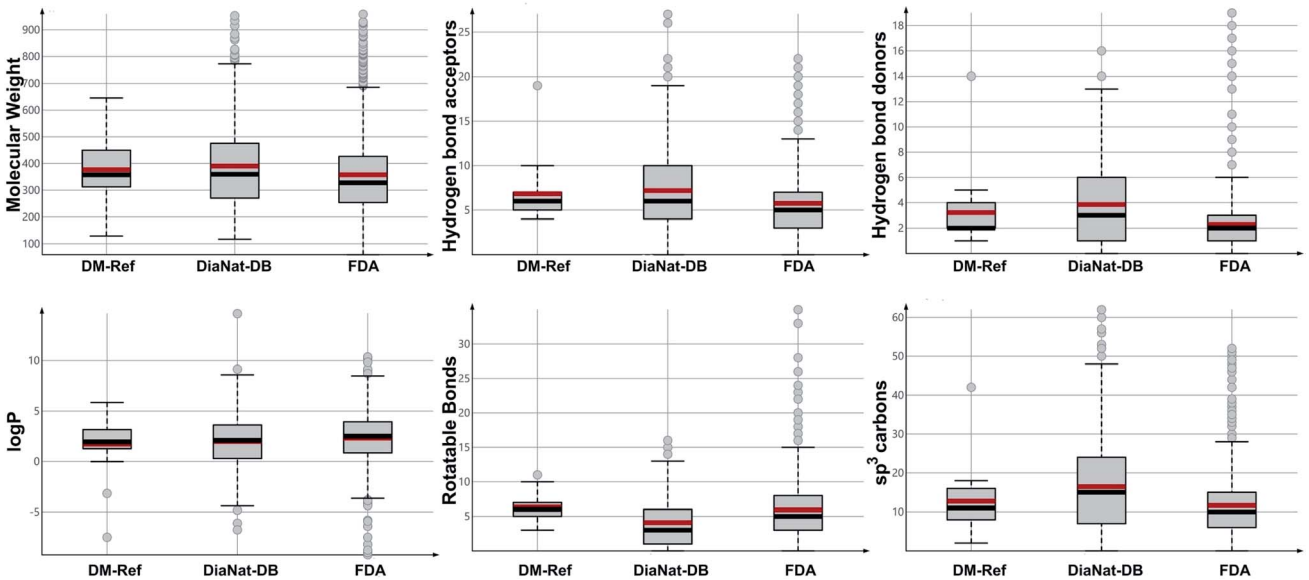

Fig. 1 Box plots of physicochemical descriptors based on drug-like properties for the three databases analyzed. Outliers are shown as gray circles, the gray box represents the interquartile range, and the median and mean values are depicted as red and black lines inside the boxes, respectively.

3.3. Taxonomic groups represented in DiaNat-DB

DiaNat-DB contains information about 211 plant species with antidiabetic activity. These plants belong to 91 botanical families. Based on the number of genera for each family and classified based on the use of the plant (edible or medicinal), the most represented families are *Fabaceae*, *Myrtaceae*, *Asteraceae*, *Moraceae*, and *Rutaceae*, having more than ten compounds each (Fig. S1†).

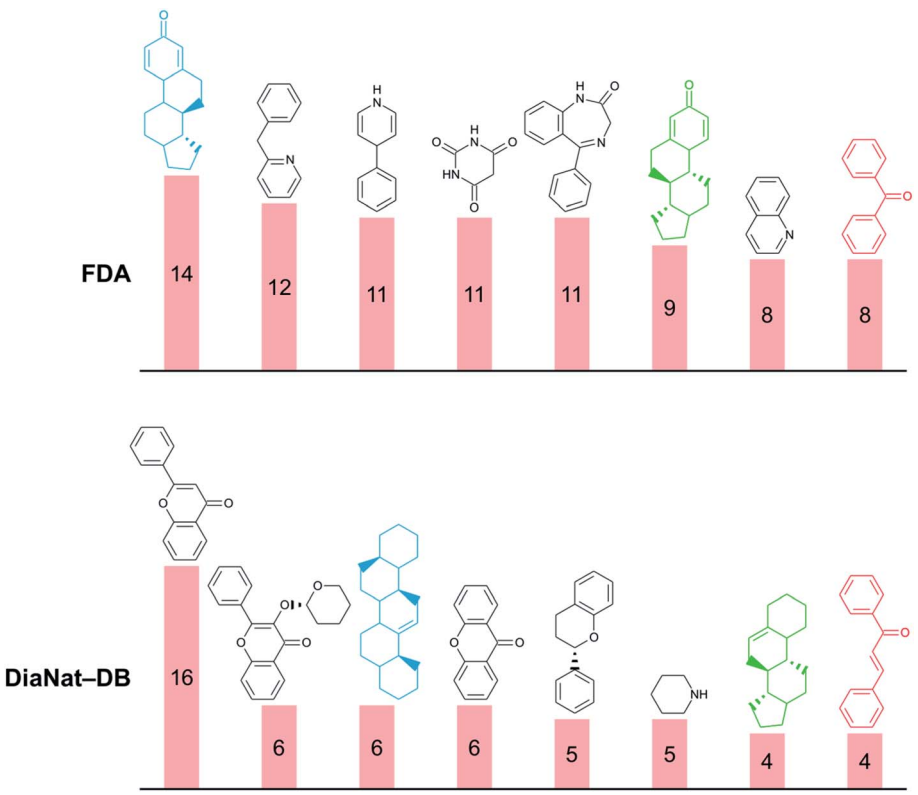

Fig. 2 The eight most frequent scaffolds in FDA and DiaNat-DB. Scaffolds with the same color (blue, green, or red) have a Tanimoto similarity > 0.7. Benzene was the most common Bemis–Murcko scaffold in both databases (143 in FDA and 17 in DiaNat-DB), and it is not included on this graph.

3.4. Profiling of physicochemical properties of DiaNat-DB and reference antidiabetic drugs

Chemoinformatic characterization of natural product databases provides a quick overview of properties and structural features and is frequently used in drug discovery campaigns.[35,36] Fig. 1 shows the distribution of six drug-like physicochemical properties of pharmaceutical interest, computed as described in the Methods section. Results indicate that compounds in DiaNat-DB are diverse regarding their physicochemical properties. On average, molecules in DiaNat-DB have slightly greater molecular weight than FDA-approved drugs ($p = 0.002$, t-test) and more hydrogen bond acceptors ($p < 0.001$, t-test) and donors ($p < 0.001$, t-test). Notably, despite that the number of rotatable bonds is on average smaller than those in the other two reference databases ($p < 0.001$, t-test), it tends towards a higher proportion of sp^3 carbon atoms ($p < 0.001$, t-test), indicating greater molecular complexity.

3.5. Scaffold content analysis and similarity with antidiabetic drugs

Fig. 2 shows the most frequent molecular scaffolds in FDA and DiaNat-DB. The benzene scaffold that is the most frequently found in several compound databases of small molecules[37,38,40] is omitted in the figure. DiaNat-DB and FDA contain 963 Bemis–Murcko scaffolds in total. The most frequent scaffold in DiaNat-DB is 2-cyclohexyl-4*H*-chromen-4-one (16 compounds with this scaffold). We also found compounds and scaffolds from DiaNat-DB that are very similar (Tanimoto similarity > 0.7) to small molecules in the FDA database (see Tables S1 and S2 in the ESI†).

3.6. Chemical space and hit expansion

A total of 224 molecules in DiaNat-DB had at least one analog in 445 cores found both in ZINC15 and DiaNat-DB. We could reduce the number of cores to 333 by discarding redundant

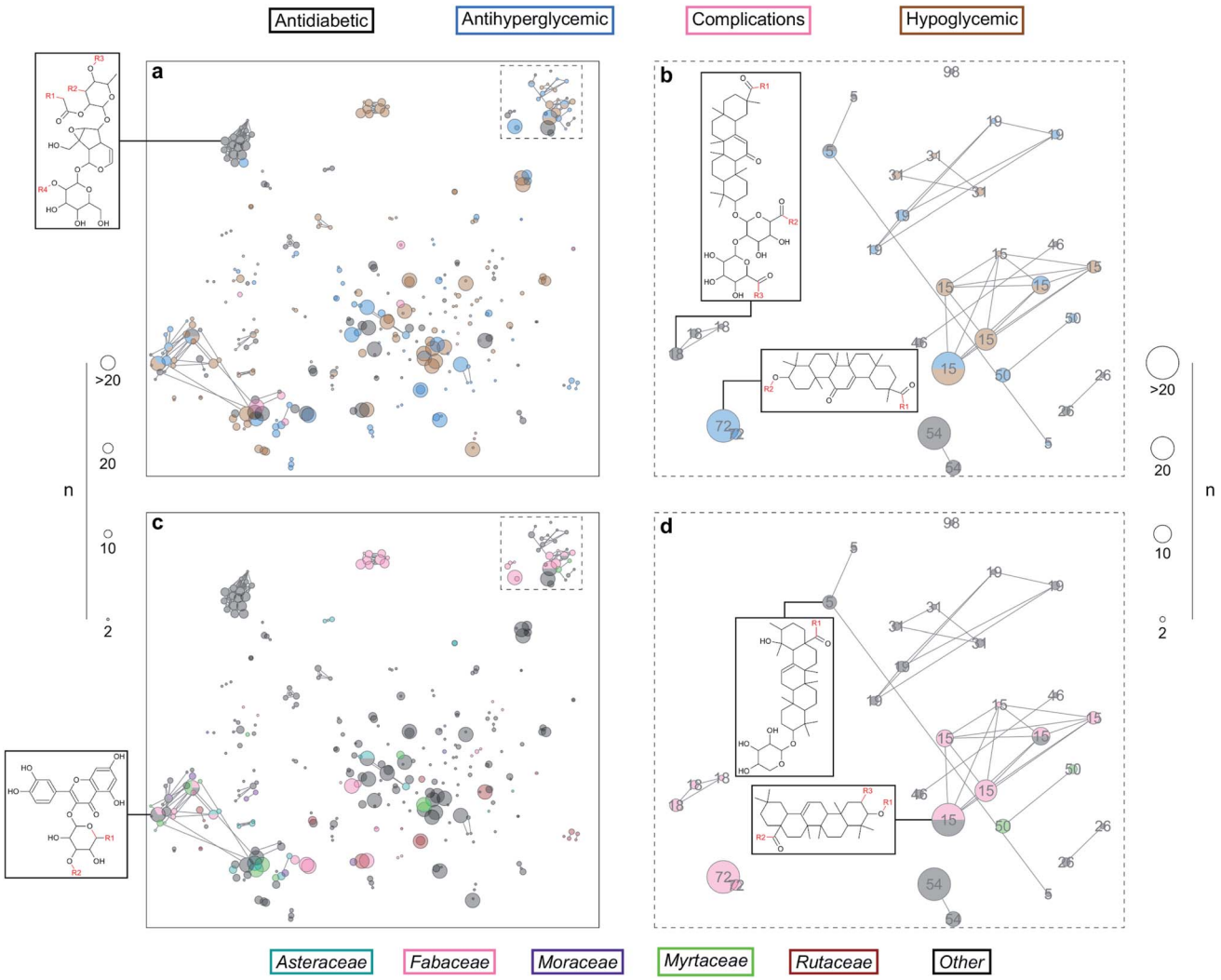

Fig. 3 Constellation plots of DiaNat-DB + analogs from ZINC15. Every dot represents one of the 333 cores represented both in DiaNat-DB and ZINC15. The size of the dot indicates the number of molecules mapped to the core. Dots are colored according to the information regarding the compounds in DiaNat-DB: their activity in (a) and (b), and the family of the plants they were extracted from in (c) and (d). Panels (b) and (d) are a "zoom-in" of the area indicated with dashed lines, where dots are annotated by the ID of the analog series they belong to.

cores matching the same compounds as any other core with more heavy atoms. In total, the ZINC15 analogs of compounds in DiaNat-DB comprise 8356 unique structures. Fig. 3 shows a constellation plot – a chemical space representation of all the analog series found. We effectively expanded the chemical space of DiaNat-DB by a factor of 24 with this methodology. The found analogs can be further tested to increase the knowledge of the structure–activity relationships for these compounds. In addition, filters based on the prediction of toxicological endpoints trims down the number of molecules for further development. The most common predicted toxicological endpoints include mutagenicity, teratogenicity and acute toxicity;[42] applied to metabolites and agrochemicals for regulatory affairs.[41] ZINC15 analogs of molecules in DiaNat-DB and comprehensive R tables for all cores are available in the ESI.†

4. Conclusions

We introduce the first version of a novel and unique antidiabetic, natural product database called DiaNat-DB. The current (first) version has 336 compounds. A property and structural analysis of the database indicated that molecules in DiaNat-DB have, in general, drug-like properties. Additionally, we report scaffolds that are unique to DiaNat-DB. The correlation of the most frequent chemotypes with biological activity allowed to establish which structural scaffolds are present in compounds with antihyperglycemic and hypoglycemic activities. The broad structural diversity and complexity of compounds in DiaNat-DB suggest its potential to yield molecules with high target selectivity in biochemical assays. Finally, the hit expansion analysis allowed to expand from 336 compounds into more than eight thousand analogs and represent starting points of lead-optimization programs inspired by natural products. DiaNat-DB is publicly available. DiaNat-DB represents a step forward to the integration of chemcial and biological information for the development of antidiabetic compounds from natural origin. Indeed compound databases are a cornerstone in chemoinformatics and other informatics-related disciplines that have made key contributions to chemistry, biology, and biomedical sciences.[39]

Author contributions

AMM and KMM conceptualized the work, wrote the manuscript, and discussed the results. KNC and JJN prepared the database and performed the analysis; JJN and JLMF contributed to reviewing, editing, and discussing the manuscript.

Conflicts of interest

The authors declare no conflicts of interest.

Acknowledgements

This work was supported by Instituto de Química-UNAM, CONACyT (project numbers 286854, 220392), and DGAPA-UNAM (PAPIIT IN210518). JJN thanks CONACYT for a postdoctoral scholarship. The authors thank ChemAxon for kindly providing academic licenses of their software.

References

1 American Diabetes Association, Diagnosis and classification of diabetes mellitus, *Diabetes Care*, 2010, **33**(suppl 1), S62–S69.
2 Y. Zheng, S. H. Ley and F. B. Hu, Global aetiology and epidemiology of type 2 diabetes mellitus and its complications, *Nat. Rev. Endocrinol.*, 2018, **14**(2), 88–98.
3 IDF Diabetes Atlas 9th edition 2019 [Internet], cited 2020 Nov 17, https://diabetesatlas.org/en/.
4 J. M. Forbes and M. E. Cooper, Mechanisms of diabetic complications, *Physiol. Rev.*, 2013, **93**(1), 137–188.
5 T. T. W. van Herpt, R. F. H. Lemmers, M. van Hoek, J. G. Langendonk, R. J. Erdtsieck, B. Bravenboer, *et al.* Introduction of the DiaGene study: clinical characteristics, pathophysiology and determinants of vascular complications of type 2 diabetes, *Diabetol. Metab. Syndr.*, 2017, **9**, 47.
6 A. Y. Y. Cheng and I. G. Fantus, Oral antihyperglycemic therapy for type 2 diabetes mellitus, *CMAJ*, 2005, **172**(2), 213–226.
7 F. D. Yakaryılmaz and Z. A. Öztürk, Treatment of type 2 diabetes mellitus in the elderly, *World J. Diabetes*, 2017, **8**(6), 278–285.
8 B. Claggett, J. M. Lachin, S. Hantel, D. Fitchett, S. E. Inzucchi, H. J. Woerle, *et al.* Long-Term Benefit of Empagliflozin on Life Expectancy in Patients With Type 2 Diabetes Mellitus and Established Cardiovascular Disease, *Circulation*, 2018 Oct 9, **138**(15), 1599–1601.
9 S. Loraine and J. A. Mendoza-Espinoza, *Las plantas medicinales en la lucha contra el cáncer, relevancia para México*, Revista Mexicana de Ciencias Farmacéuticas, 2010.
10 V. Bankova, Chemical diversity of propolis and the problem of standardization, *J. Ethnopharmacol.*, 2005, **100**(1–2), 114–117.
11 D. J. Newman and G. M. Cragg, Natural Products as Sources of New Drugs over the Nearly Four Decades from 01/1981 to 09/2019, *J. Nat. Prod.*, 2020, **83**(3), 770–803.
12 M. Eddouks, D. Chattopadhyay, V. De Feo and W. C. Cho, Medicinal plants in the prevention and treatment of chronic diseases, *J. Evidence-Based Complementary Altern. Med.*, 2012, **2012**, 458274.
13 P. K. Mukherjee, K. Maiti, K. Mukherjee and P. J. Houghton, Leads from Indian medicinal plants with hypoglycemic potentials, *J. Ethnopharmacol.*, 2006, **106**(1), 1–28.
14 S. M. Escandón-Rivera, R. Mata and A. Andrade-Cetto, Molecules Isolated from Mexican Hypoglycemic Plants: A Review, *Molecules*, 2020, **25**(18), 4145.
15 A. Madariaga-Mazón, R. B. Hernández-Alvarado, K. O. Noriega-Colima, A. Osnaya-Hernández and K. Martinez-Mayorga, Toxicity of secondary metabolites, *Phys. Sci. Rev.*, 2019, **4**(12), 20180116.
16 P. A. Meetei, P. Singh, P. Nongdam, N. P. Prabhu, R. Rathore and V. Vindal, NeMedPlant: a database of therapeutic

applications and chemical constituents of medicinal plants from north-east region of India, *Bioinformation*, 2012, **8**(4), 209–211.

17 H. Pérez-Sánchez, H. den-Haan, J. Peña-García, J. Lozano-Sánchez, M. E. Martínez Moreno, A. Sánchez-Pérez, *et al.* DIA-DB: A Database and Web Server for the Prediction of Diabetes Drugs, *J. Chem. Inf. Model.*, 2020, **60**(9), 4124–4130.

18 A. Khatoon, I. Rashid, S. Shaikh, S. M. D. Rizvi, S. Shakil, N. Pathak, *et al.* ADNCD: a compendious database on anti-diabetic natural compounds focusing on mechanism of action, *3 Biotech*, 2018, **8**(8), 361.

19 M. U. Makheswari and D. Sudarsanam, Database on Antidiabetic indigenous plants of Tamil Nadhu, India, *Int. J. Pharm. Sci. Rev. Res.*, 2012, **3**(2), 287–293.

20 D. Weininger, SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules, *J. Chem. Inf. Model.*, 1988, **28**(1), 31–36.

21 D. S. Wishart, Y. D. Feunang, A. C. Guo, E. J. Lo, A. Marcu, J. R. Grant, *et al.* DrugBank 5.0: a major update to the DrugBank database for 2018, *Nucleic Acids Res.*, 2018, **46**(D1), D1074–D1082.

22 C. A. Lipinski, F. Lombardo, B. W. Dominy and P. J. Feeney, Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings, *Adv. Drug Delivery Rev.*, 2001, **46**(1–3), 3–26.

23 D. F. Veber, S. R. Johnson, H.-Y. Cheng, B. R. Smith, K. W. Ward and K. D. Kopple, Molecular properties that influence the oral bioavailability of drug candidates, *J. Med. Chem.*, 2002, **45**(12), 2615–2623.

24 O. Méndez-Lucio and J. L. Medina-Franco, The many roles of molecular complexity in drug discovery, *Drug Discovery Today*, 2017, **22**(1), 120–126.

25 W. Wei, S. Cherukupalli, L. Jing, X. Liu and P. Zhan, Fsp3: A new parameter for drug-likeness, *Drug Discovery Today*, 2020, **25**, 1839–1845.

26 G. W. Bemis and M. A. Murcko, The properties of known drugs. 1. Molecular frameworks, *J. Med. Chem.*, 1996, **39**(15), 2887–2893.

27 L. Maaten and G. Hinton, Visualizing Data using t-SNE, *J. Mach. Learn. Res.*, 2008, **9**, 2579–2605.

28 J. J. Naveja and J. L. Medina-Franco, Consistent Cell-selective Analog Series as Constellation Luminaries in Chemical Space, *Mol. Inf.*, 2020, 2000061.

29 J. J. Naveja, M. Vogt, D. Stumpfe, J. L. Medina-Franco and J. Bajorath, Systematic Extraction of Analogue Series from Large Compound Collections Using a New Computational Compound-Core Relationship Method, *ACS Omega*, 2019, **4**(1), 1027–1032.

30 J. J. Naveja, B. A. Pilón-Jiménez, J. Bajorath and J. L. Medina-Franco, A general approach for retrosynthetic molecular core analysis, *J. Cheminf.*, 2019, **11**(1), 61.

31 J. J. Naveja and J. L. Medina-Franco, Finding constellations in chemical space through core analysis, *Front. Chem.*, 2019, **7**, 510.

32 T. Sterling and J. J. Irwin, ZINC 15–Ligand Discovery for Everyone, *J. Chem. Inf. Model.*, 2015, **55**(11), 2324–2337.

33 Y. Chen, C. de Bruyn Kops and J. Kirchmair, Data Resources for the Computer-Guided Discovery of Bioactive Natural Products, *J. Chem. Inf. Model.*, 2017, **57**(9), 2099–2111.

34 J. L. Medina-Franco, Towards a unified Latin American Natural Products Database: LANaPD, *Future Sci. OA*, 2020, **6**(8), FSO468.

35 Y. Chen and J. Kirchmair, Cheminformatics in Natural Product-Based Drug Discovery, *Mol. Inf.*, 2020, 2000171.

36 K. Martinez-Mayorga, A. Madariaga-Mazon, J. L. Medina-Franco and G. Maggiora, The impact of chemoinformatics on drug discovery in the pharmaceutical industry, *Expert Opin. Drug Discovery*, 2020, **15**(3), 293–306.

37 T. D. Tran, S. M. Ogbourne, P. R. Brooks, N. Sánchez-Cruz, J. L. Medina-Franco and R. J. Quinn, Lessons from Exploring Chemical Space and Chemical Diversity of Propolis Components, *Int. J. Mol. Sci.*, 2020, **21**(14), 4988.

38 A. H. Al Sharie, T. El-Elimat, Y. O. Al Zu'bi, A. J. Aleshawi and J. L. Medina-Franco, Chemical space and diversity of seaweed metabolite database (SWMD): A cheminformatics study, *J. Mol. Graphics Modell.*, 2020, 107702.

39 E. López-López, J. Bajorath and J. L. Medina-Franco, Informatics for chemistry, biology, and biomedical sciences, *J. Chem. Inf. Model.*, 2020, DOI: 10.1021/acs.jcim.0c01301.

40 K. Martinez-Mayorga, Characterization of a comprehensive flavor database, *J. Chemom.*, 2011, **25**, 550–560.

41 G. Gómez-Jiménez, K. Gonzalez-Ponce, D. J. Castillo-Pazos, A. Madariaga-Mazon, J. Barroso-Flores, F. Cortes-Guzman and K. Martinez-Mayorga, The OECD Principles for (Q)SAR Models in the Context of Knowledge Discovery in Databases (KDD), *Adv. Protein Chem. Struct. Biol.*, 2018, **113**, 85–117.

42 R. B. Hernández-Alvarado, A. Madariaga-Mazón and K. Martinez-Mayorga, Prediction of toxicity of secondary metabolites, *Phys. Sci. Rev.*, 2019, **4**(11), 20180107.