PHCD: a database of chemical compositions of Persian medicinal herbs

Mohammad Hossein Karimi-Jafari[1], Rohoullah Firouzi[2,*], Mitra Ashouri[3], Atefeh Poursoleiman[1]

[1]Department of Bioinformatics, Institute of Biochemistry and Biophysics, University of Tehran, Tehran, Iran

[2]Department of Physical Chemistry, Chemistry and Chemical Engineering Research of Iran, Tehran, Iran

[3]Department of Physical Chemistry, School of Chemistry, College of Science, University of Tehran, Tehran, Iran

*To whom correspondence should be addressed.

Abstract

Motivation: Herbal compounds exhibit a rich source of diverse molecular scaffolds with a high level of structural complexity that is crucial for *in silico* and *in vitro* screening in the drug discovery process.

Results: The Persian Herbal Constituents Database (PHCD) has been designed as a free webserver for providing useful information on many famous Persian medicinal herbs and their phytochemical ingredients. The PHCD facilitate in-depth plant and chemical searches by name, physicochemical properties, substructure, and molecular similarity using multiple advanced search functions.

Availability: PHCD is freely available at: https://persianherb.com.

Contact: rfirouzi@ccerci.ac.ir and firouzi.chemist@yahoo.com

1 Introduction

Medicinal plants have always played an important role in maintaining and improving human health for thousands of years. They have been widely used in traditional medicines and are still served as a valuable natural source for modern drug discovery (Rodrigues et al., 2016). To facilitate access to phytochemical and pharmacological information on medicinal plants, and consequently to accelerate the drug design process, many useful herbal compound databases have been established in recent years (Lagunin et al., 2014). Most of these databases cover the medicinal plants of specific geographical regions and therefore, their informational content is relatively different from each other. These differences are due to the fact that some plants are

endemic and only found in certain locations on the planet and also, lists of constituent compounds present in the plants are highly dependent on climate conditions which means that the herbal ingredients of the same plant may vary in different regions of Earth. In this light, new published databases for well-characterized phytoconstituents of the medicinal plants from different regions of the world have become a necessity for the development of new drug candidates (Nicola et al., 2012). Here, we present a new searchable database containing useful information about a large number of famous Persian medicinal herbs and their chemical constituents. Persia is located within the Middle East region – a transcontinental zone between Asia, Africa, and Europe – with diverse climatic conditions and high variety of plant species, which is absent from all currently available databases.

2   Data description and preparation

This database contains two data sets: herb information and compound information. The general information such as scientific name, common name, Persian name, family, medicinal use(s), and portrait(s) for each herb was gathered from published books and web resources. The lists of phytochemical ingredients found in these herbs were manually extracted from only peer-reviewed scientific publications. The initial 3D structures of all collected compounds were either downloaded from the PubChem database or were drawn and edited using GaussView0.5 or HyperChem7 programs based on the 2D structures presented in the articles reviewed. The obtained geometries were structurally optimized at the dispersion-corrected semi-empirical method PM7 with the MOPAC2016 program (Stewart, 2016). All optimized compounds were annotated by various representations including 1D strings such as SMILES (generated with the OpenBabel program (O'Boyle et al., 2011)), InChI, InChIKey (generated by the InChI Software), and CDK fingerprint (produced by the PaDEL software (Yap, 2011)). The compounds were also annotated by molecular properties. These properties include some electronic features (computed by the MOPAC2016) such as HOMO-LUMO gap, electric dipole moment, ionization potential and also, several important physicochemical descriptors (calculated by the PaDEL) such as molecular weight, molecular volume, XlogP, number of rotatable bonds, number of hydrogen bond acceptors and donors, which are often used for predicting "drug-likeness" and oral bioavailability of small molecules.

3 Data access and search methods

The data in PHCD can be easily accessed and mined by a variety of search types and filters. The screenshots of the search interfaces and result pages have been demonstrated in Fig. 1. The 'Basic search' (Fig. 1A) and the 'Quick search' (Fig. 1B, the search box at the PHCD homepage) enable one to find desired herb(s) or compound by just typing name or other text annotations. The 'Physicochemical properties search' (Fig. 1C) makes it possible to select a subset of compounds by specifying ranges for any combination of some important physicochemical descriptors calculated which are often used for predicting "drug-likeness" of compounds. The 'Topological search' (Fig. 1D) provides a flexible query containing the count of particular atoms, bonds, rings, and fused rings, for finding compound(s) with specific molecular formula and topological features. Finally, the 'Structural similarity search' (Fig. 1E) has been implemented in the PHCD, based on binary fingerprints and the Tanimoto coefficients. In this search, one can draw or import a molecular structure and compare it with compounds of the PHCD database. Therefore, by using this advanced search, structural similarities between the pool of compounds present in the PHCD and FDA-approved drugs, human metabolites, and other available sets of ligands/hits/leads compounds can be searched and analyzed to find alternative herbal compounds for the drugs and metabolites and also, identify the mechanisms of action or possible targets of the herbal compounds. In addition, if one has only 2D images of some molecules and does not know their exact names, the similarity search will be a useful tool for finding these molecules or their similar molecules in the PHCD.

Dependent on the above searching methods, one or more herbs or compounds will be returned to users, as shown in Fig. 1. By clicking on each record of the list of herbs or compounds, all information related to the herb or the compound is displayed. The results page for each herb (Fig. 1X) or compound (Fig. 1Y) is separated into different sections that a click on the drop-down arrow of each section reveals detailed information on the herb or the compound. These results pages link directly to each other by clicking on the names in the herb or the compound list on the pages (dashed and dotted-lines in Fig.1). Each results page contains links to the first page of articles from which the herbal ingredients have been extracted and it has been also provided links to other useful online databases to obtain further information – Wikipedia, GBIF, and DuKe's database for herbs and PubChem for compounds. All the information calculated for each compound can be downloaded in .sdf or .xml format.
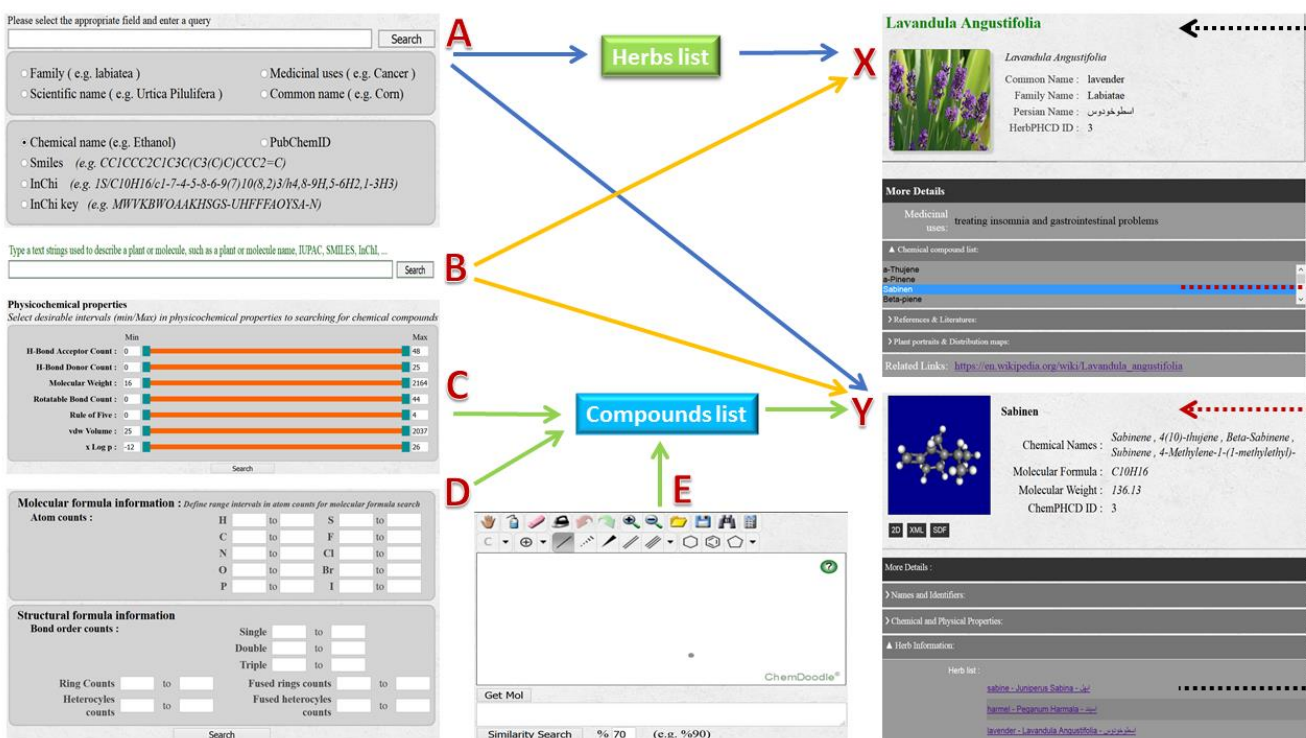
Fig. 1. Screenshots of the search interfaces, the result pages, and links between them.

4 Web-based database implementation

PHCD is freely available at: http://persianherb.com. It is constructed on the SQL Server 2013 as a relational database management system. The Web interface was designed using HTML, CSS, and .NET Framework 4 and also the JME molecular editor is used for structure drawing in the structural similarity search.

Funding

Conflict of Interest: none declared.

References

Lagunin,A.A. *et al*. (2014) Chemo- and bioinformatics resources for in silico drug discovery from medicinal plants beyond their traditional use: a critical review. *Nat. Prod. Rep.,* **31**, 1585−1611.

Nicola,G. *et al*. (2012) Public Domain Databases for Medicinal Chemistry. *J. Med. Chem.* **55**, 6987−7002.

O'Boyle,N.M. *et al*. (2011) Open Babel: an open chemical toolbox. *J. Cheminform.* **3**, 33−47.

Rodrigues,T. *et al*. (2016) Counting on natural products for drug design. *Nature Chem.* **8**, 531−541.

Stewart,J.J.P. (2016) MOPAC2016; Stewart Computational Chemistry: Colorado Springs, CO.

Yap,C.W. (2011) PaDEL-Descriptor: An open source software to calculate molecular descriptors and fingerprints. *J. Comput. Chem.* **32**, 1466−1474.