UNaProd: A Universal Natural Product Database for *Materia Medica* of Iranian Traditional Medicine

Ayeh Naghizadeh[1†], Donya Hamzeheian[1†], Jehad Aldahdooh[2†], Shaghayegh Akbari[1], Fahimeh Mohammadi[3], Tohid Otoufat[3], Saeme Asgari[4], Azadeh Zarei[1], Samane Noroozi[1], Najmeh Nasiri[5], Mahdi Salamat[1], Reza Karbalaei[6], Mehdi Mirzaei[7], Hossein Rezaeizadeh[1], Jing Tang[2,8], Mehrdad Karimi[1], Mohieddin Jafari[2]

[1]Department of Traditional Medicine, School of Persian Medicine, Tehran University of Medical Sciences, Tehran, Iran.
[2]Institute for Molecular Medicine Finland (FIMM), Helsinki Institute of Life Science, University of Helsinki, Helsinki, Finland.
[3]Department of Chemistry, K. N. Toosi University of Technology, Tehran, Iran.
[4]Medical Biotechnology Department, Biotechnology Research Center, Pasteur Institute of Iran, Tehran, Iran.
[5]Genetics and Agricultural Biotechnology Institute of Tabarestan (GABIT), Sari Agricultural Sciences and Natural Resources University (SANRU), Sari, Iran.
[6]Department of Biology, Temple University, PA, USA.
[7]Department of Applied Mathematics, Faculty of Mathematical Sciences, Tarbiat Modares University, Tehran, Iran.
[8]Research Program in Systems Oncology, Faculty of Medicine, University of Helsinki, Helsinki, Finland

[†] Those authors were contributed equally as first author.

*Corresponding authors:

Mohieddin Jafari, Ph.D.
Mohieddin.jafari@helsinki.fi

Mehrdad Karimi, MD, Ph.D.
mehrdadkarimi@yahoo.com

Abstract

As a holistic medical school, Iranian traditional medicine (ITM) considers the human body as a dynamic and intricate network of interconnecting processes. Currently, systems biology and more precisely systems medicine and pharmacology can be an aid in providing rationalizations for many traditional medications and treatments and elucidating a great deal of knowledge they can offer to guide future research in medicine. Therefore, re-organization and standardization of traditional medicine data are requested more than ever before. To address this issue, we have constructed UNaProd, a Universal Natural Product database for *materia medica* of ITM. Primarily based on *Makhzan al-Advieh*, which is the most recent encyclopedia of *materia medica* in ITM with the largest number of monographs, this database was created using both text mining methods and manual editing. UNaProd is currently hosting to 2696 monographs from herbal to animal to mineral compounds in 16 diverse attributes such as origin and scientific name. In the current version, UNaProd is hyperlinked to IrGO and CMAUP databases for *Mizaj* and molecular features respectively and it is freely available at http://jafarilab.com/unaprod/.

Introduction

Iranian Traditional Medicine

Complementary and alternative medicine (CAM) refers to practices of traditional or modern systems that are not considered part of mainstream medicine (1). Traditional systems of healing, including Chinese Traditional Medicine (TCM), Ayurveda, Unani medicine and Iranian Traditional Medicine (ITM) or so-called Persian Medicine have a theoretical foundation and embrace the indigenous knowledge, skills and practices of ancient civilizations to maintain health, prevent disease, and treat physical or mental illness. The need for an integrative approach to use both complementary and conventional health care is increasingly recognized by many countries and allows safe and effective access to traditional medicines (2). A medical system notable in terms of theory and clinical experience, ITM is founded on philosophical concepts associated with the four elements (3) (e.g. see Table 1). Practiced for thousands of years, ITM is distinctive for world-renowned polymaths such as Rhazes (4) whose book *al-Hawi (Continiens)* is an encyclopedia of clinical medicine, embracing views of predecessor physicians from various countries; and Avicenna whose book, *Canon of Medicine* was a main medical reference in Europe until the 16th century (5). Proficient Iranian physicians continued to revise and complement the body of collected knowledge up to the modern period. An exemplar of the 18th century, Mohammad Hosein Aghili is recognized for books on all scopes of ITM including *Kholasat al-Hekmah* (on principles of Iranian medicine), *Mo'alejaat* (on diseases and their treatment), *Makhzan al-Advieh* (on monotherapy), and *Ghrabadin-e Kabir* (on combination therapy).

Though forsaken for nearly a century due to adoption of allopathic medicine, the academic society has recommened research in the CAM recently and integrated it into the healthcare system (6). ITM takes a personalized approach to health maintenance and healing, considering environmental and patient factors as well as signs and symptoms for disease diagnosis and treatment. Enjoying a holistic view, ITM observes the human body as an integrated system and emphasizes the interconnections between the organs. ITM physicians would consider not just the disease, but more notably the patient, to maintain or restore the balance in the body by a number of treatment modalities including lifestyle modification, drugs, and procedures.

The holistic view in medicine offered by traditional medical schools gradually diminished in the late twentieth century as a result of the emergence of molecular biology, which focuses on the molecular basis of biological mechanisms. Despite all the advances and elucidation of a vast amount of molecular mechanisms, one-disease one-target paradigm has its limits and new trends of pharmacology such as polypharmacology and drug repurposing have developed to cure complex diseases which can rarely be due to an impairment or dysfunction in just one gene or molecule (7). Recently, the deficiencies of the dominant reductionism in medical research have become more evident, resulting in attention towards systems biology to understand the complex interconnections within biological systems (8-19).

While a system can be defined as a unified whole comprised of regularly interacting or interdependent components (14), systems biology is an attempt to illuminate the whole

3

framework of biological processes by considering all components in the biological entities as a system rather than individually. It facilitates a deeper understanding of the physiologic and pathologic processes, as well as recognizing the right targets to treat diseases (17, 20). Development of systems medicine and omics technologies have been an aid in providing rationalizations for traditional medicines, proving their efficacy and illuminating the great deal of knowledge they can offer to guide future research in medicine (21-23).

On the other hand, there are several methods to discover new pharmacologically active plant compounds including biodiversity-based (24), chemosystematics (25), ecological (26), computational (27), and ethnopharmacological (28) approaches. Exploiting ethnomedical data, which includes both folklore and organized traditional medical systems, is a time and cost-effective method to discover new pharmacological compounds with higher hit-rate compared to biodiversity-based methods (29, 30). In fact, databases of traditional medicine can be a valuable resource for new drug discovery. For example, TCM scientists have addressed this issue and developed a number of databases such as Traditional Chinese Medicine Integrated Database (TCMID) (31, 32) or Collective Molecular Activities of Useful Plants (CMAUP) (33), which are host to data on all aspects of TCM, including diseases, drugs and their targets. These databases also provide external links to other resources including OMIM (Online Mendelian Inheritance in Man), Drugbank, PubChem, and STITCH (search tool for interactions of chemicals) (31).

However, there is lack of ITM databases available to the academic society. An apt starting point would be creating a database of *materia medica,* one of the key features in traditional medical systems including ITM. The body of collected knowledge on the therapeutic characteristics of herbal, animal and mineral compounds has been extensively documented and revised by ITM physicians throughout the millennia (34). Properties of medicinal substances as described in formal medical literature, including ITM resources are deducted from the foundational principles of that medical system as well as clinical experience of physicians. In order to exploit the full potential of traditional medical resources, their data and concepts should be re-organized and standardized to create a shared interface with conventional medicine.

To address the data integration problem, we created Universal Natural Products Resource (UNaProd), a database of medicinal substances consisting of 2696 drug monographs based on an encyclopedia of ITM *materia medica* i.e. *Makhzan al-Advieh*, a foundation on which more comprehensive databases of ITM can be built. This database will help exploit knowledge from ITM resources in a number of ways. Firstly, ITM can be expressed in a terminology that can be used by modern medicine. Second, the relationships between ITM concepts will be more precisely determined, which will in turn unveil the underlying biological system of this medical school. Third, providing justifications for traditional use of plants by linking them to their associated compounds will facilitate integration of ITM with both allopathic medicine, and other traditional medical systems including TCM. Fourth, it will provide an extensive range of options for discovery of new drugs. As a time- and money-consuming process, new drug discovery will indeed be

facilitated through utilizing the connections ITM has established between a medicinal substance and a specific disease.

Table 1: Some basic Definitions of ITM

Term	Definition
Qualities	The four qualities of hotness, coldness, wetness, and dryness are foundational concepts in ITM, by which phenomena including the effects of medicinal substances on the body are explained.
Active quality	Hotness and coldness are considered effectual on wetness and dryness, and thus called active qualities.
Passive quality	Wetness and dryness are considered affectable by hotness and coldness, and thus called passive qualities.
Mizaj	*Mizaj* is defined as the uniform quality, created via interaction of the four qualities constituting a substance in various proportions. A prediction of the effects of medicinal substances on the body is possible via this property.
Mizaj type	Described as either balanced or unbalanced and expressed as exclusively one of the qualities or a combination of an active and a passive quality.
Mizaj degree	The potency of each of the qualities in a *Mizaj* type is mostly expressed in a range between one to four (least to most potent), each further classified into minimum, medium, or maximum.
DNS (Degree Not Specified)	In UNaProd database, DNS is ascribed to qualities of *Mizaj* for which a *Mizaj* type has been determined, while no degree has been specified.

Materials and methods

UNaProd database was created by means of both text mining and manual editing by an ITM committee comprised of specialists and researchers in the Department of Persian medicine, Tehran University of Medical Sciences. The construction pipeline is illustrated in Fig. 1.
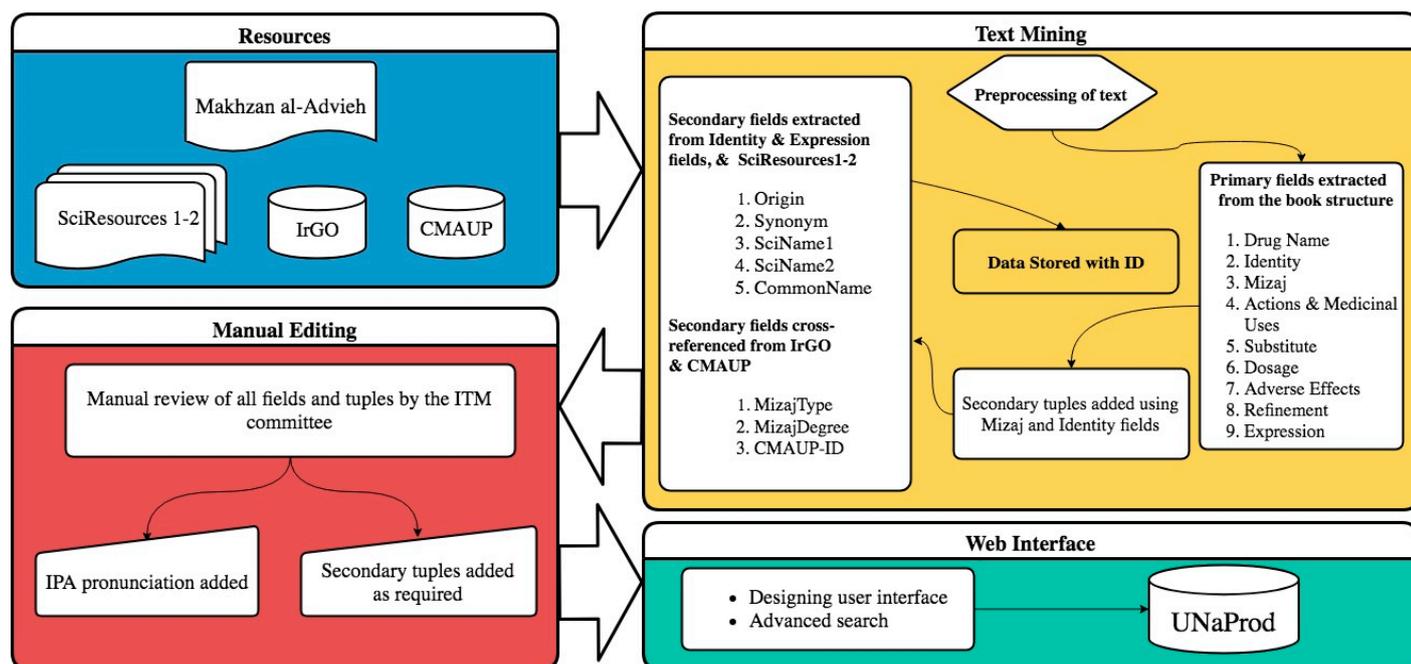

Figure 1: UNaProd construction pipeline. The database was created by text mining Makhzan al-Advieh to extract the primary fields from the text, and secondary fields from SciResources, IrGO, and CMAUP. Subsequently, all fields were manually reviewed and secondary tuples were added to the database. In the next step, the web interface was designed to create an online database.

Resources

Among the resources available in ITM concerning documentations of medicinal substances, *Makhzan al-Advieh* written by Mohammad Hosein Aghili in 1769 A.D. is the most recent encyclopedia of *materia medica* and contains the greatest number of drug monographs; thus, it was selected as the resource for the raw data of our database. The *Makhzan al-Advieh* manuscript consists of an introduction describing the contents of the book and thirteen chapters, and an appendix providing synonyms and translations. Twelve chapters are on preliminary topics including term definitions, estimation of drug *Mizaj*, units of weight, instructions on drug preparations, and astronomical considerations. The main chapter of the book is comprised of drug monograph documentation in alphabetical order. Drug monographs are described in terms of synonyms, identity, *Mizaj*, actions and medicinal uses, dosage, adverse effects, refinement, and substitute in this semi-structured book.

Furthermore, Aghili has mentioned viewpoints of other physicians and pharmacologists in the cases where there is no consensus on a certain characteristic, or when prominent scholars like *Avicenna*, *Galen* and *Ibn-e Baitar* have other opinions.

The initial part of each drug monograph description includes a description of its correct pronunciation, synonyms, and translations. About 70 languages and dialects have been used to provide translations, the most common of which include Persian, Arabic, Indian and Greek. A description of the identity of each drug monograph in terms of its origin, types and morphological characteristics is provided in the second part. Most drug monographs are of either herbal, animal or mineral origin, but there are also instances of compound drug monograph like *Maligharaten*, which is honey boiled in water. Included in this section are the following: types and varieties of the medicinal substance, for example being wild or cultivated for plants and wild or domestic for animals; the morphological characteristics of drug parts such as leaves, bark, flower or fruit; the place of origin including countries and geographical locations where the drug monograph may be found. In the cases where the identity of a drug is in dispute, Aghili has mentioned views of different scholars. The third part of each drug monograph is dedicated to *Mizaj*, which is a key concept in Iranian traditional medicine and an attribute of all that exists in the universe. Persian scholars determined the *Mizaj* of a medicinal substance through deduction and experience (35). Subsequently discussed are the actions and medicinal uses, which encompass a list of pharmacological actions followed by details on the medical conditions which benefit from the drug. Adverse effects, refinements, dosage and substitute comprise the succeeding parts of drug monograph documentation. Adverse effects include the body organs or individuals that may be harmed by the drug, and is usually followed by refinement, which includes particular drug preparation or another drug to reduce or eliminate the mentioned side effects. Dosage has been provided for some of the drugs, and for more potent ones the lethal dose has also been specified. Also included in some instances is substitute defined as a drug with similar effects that can be used as a replacement. The potency of the substitute drug has been reported as the relative proportion that should be prescribed to reach a similar effect.

Two resources were used to extract scientific names of the herbal substances. Composed by *Ghahreman* and *Okhovvat*, the first resource (*'SciResource1'*) titled 'Matching the Old Medicinal Plant Names with Scientific Terminology', includes a history of medicinal plants, the methodology of using ancient texts on medicinal plants, the old medicinal plant names, and identification of scientific names for plants (36). The second resource ('SciResource2') includes the appendix provided in the corrected edition of Makhzan al-Advieh ('MakhzanCE'), from which both scientific names and common names were extracted (35).

IrGo (Iranian traditional medicine General Ontology) (37), which currently consists of ontology terms for *Mizaj*, was used as the resource for the *Mizaj* type and degree provided in *Makhzan al-Advieh*.

Additionally, plants with scientific names existing in CMAUP were linked to UNaProd, covering a total of 5645 plants in medicinal, food, edible, agricultural and garden plant classes. CMAUP also provides drug targets (ChEMBL target classes) and their activity levels,

7

together with GO categories, KEGG categories, and ICD blocks for gene ontologies, biological pathways, and international classification of diseases respectively.

Field and Tuple Extraction

First Phase: Text mining

Text mining methods were used to generate the primary fields in UNaProd database. Initially, the text of *Makhzan al-Advieh* was pre-processed. There are a number of Arabic characters, including *Tanwin* (أ) and *Hamzah* (ئ) that are used in different formats in Persian texts; thus, these Arabic diacritics were removed and replaced by simple *Alef* (ا) or *ye* (ی) of the Persian alphabet to facilitate searching. Also, there are two unicodes for the letters *kaf* (ک) and *ye* (ی), which can be used interchangeably and thus needed to be normalized. To perform the above-mentioned tasks, Hazm package in Python was used (38). Additionally, some prefixes and suffixes including 'می' (ing) in verbs, can be written without a separator ('میگوید'), or be separated from the verb either by a space ('می گوید'), or zero-width space ('میگوید'). Since zero-width space is not routinely used by all of the books and is not possible to type in browsers, all such cases were unified and changed to a space.

Each drug and its description were separated using regular expressions and stored in the 'Drug' table. Aghili has used certain words to demarcate the beginning of each part in the drug monograph description, which were used to separate the information of each field. Following extraction of the mentioned fields from the text, the string (a sequence of characters) not pertaining to any of those keywords was labelled as the 'Expression' field of that drug, which mainly consisted of the pronunciation and synonyms. The descriptive pronunciations in the 'Expression' field were manually translated into International phonetic alphabets (IPA) and stored in the 'Pronunciation' field. The above-mentioned fields, namely 'DrugName'; 'Mizaj'; 'Identity'; 'Substitute'; 'AdverseEffects'; 'Actions&MedicinalUses'; 'Dosage'; and 'Remarks' constituted the primary fields for each drug monograph. To determine the synonyms provided for each drug monograph, a list comprising all languages, dialects, and scholar names mentioned in the text was prepared and subsequently, terms that followed the words in this list in the 'Expression' and 'Identity' fields were determined as the synonyms of drug monographs and stored in the 'Synonym' table.

Secondary fields for each tuple include 'Pronunciation', 'Synonym', 'Origin', 'CommonName', 'SciName1', 'SciName2', 'MizajType', and 'MizajDegree'. The 'Origin' field of a drug, whether 'animal', 'herbal', 'mineral', 'compound', or 'other', was determined based on the existence of related keywords in the 'Identity' field. In most cases, the authors explicitly mentioned the origin in the first sentence of this field. Otherwise, the keywords of each origin type were matched in the remaining parts of the identity of a drug. Eventually, an origin type for each drug was assigned based on the frequency of the keywords for each group, namely herbs, animals, minerals, or compounds. In other words, the origin type with the most keywords in the 'Identity' field was assigned to the drug. Fields of 'CommonName', 'SciName1', and 'SciName2' for plants were determined by text mining 'MakhzanCE', 'SciResource1', and 'SciResource2' respectively. Animal and mineral scientific and common

8

names were manually curated. Plants with scientific names were linked to the CMAUP database.

The *Mizaj* of each drug were extracted using IrGO (Iranian traditional medicine General Ontology), and stored in a separate table called 'DrugMizaj'. Currently consisting of an ontology for *Mizaj*, IrGO was used to determine the *Mizaj* type and degree of drug monographs based on the existence of the keywords in the 'Mizaj' field. For each drug with a 'Mizaj' description, one or more rows - depending on whether other scholars' views were quoted in the text - were added to 'DrugMizaj'. The resulting table contains information regarding the ID of drugs, the scholar from whom the specified *Mizaj* was quoted, '*MizajType'* and hotness, coldness, wetness, and dryness degrees collectively named '*MizajDegree*'. The extracted types and degrees were linked to the IrGO database.

Second Phase: Manual Editing

Following extraction of the above-mentioned fields, all parts of the resultant database was reviewed manually by the ITM committee and verified with the lithograph print of Makhzan al-Advieh. Regarding many drug monographs, Aghili mentions distinct characteristics for drug parts and different types, varieties or processed form of drugs. One of the examples is quince, in which distinct *Mizaj* actions and medicinal uses have been attributed to its blossom, seed, sweet fruit, sour fruit, and purée prepared from fruit; also, features of a processed form, quince oil, have been specified. Thus, in order to create a more organized and easier access to the database, some secondary tuples were extracted from the primary ones.

Web Interface

UNaProd was built using PHP 7.2.11 for server-side data processing, Javascript ECMAScript 2015 for the frontend and Plotly library 1.40.0 for the generation of the interactive visualizations while data was stored in MySQL 10.1.37. An introduction to the database, together with a manual and statistics are available in the website. Basic and advanced search options were also provided to enable query of database contents by one or more keywords in the fields selected by the user from dropdown menus.

Results

Database Structure

Based on the primary fields, UNaProd database contained 1741 drug monographs which increased to 2696 by manually creating the secondary tuples. In addition to 'Identity', 'Mizaj', 'ActionsandMedicinalUses', 'Dosage', 'AdverseEffects', 'Refinements', 'Substitute', and 'Expression', which constitute the primary fields of the database extracted from the text of *Makhzan al-Advieh*, three fields including two columns of scientific names, a column of common name were extracted from resources other than the text. An origin was ascribed to each drug based on the identity text fields (Fig. 1). *Mizaj* types and degrees of drug monographs were extracted using IrGo database and stored in a separate 'DrugMizaj' table. Furthermore, synonyms mentioned by other scholars or translations were extracted from the 'Expression' and 'Identity' fields and stored in a separate 'Synonyms' table. The diagram of the resultant database was depicted in Supplementary file 1.

Drug Origins

Drugs were defined to have herbal, animal, mineral or other origins, the percentages of which and their share in the main *Mizaj* types were shown in Fig. 2A and 2B. Herbs clearly constituted the majority of drug monographs, followed by medicinal substances with animal and mineral origin that comprise approximately 17% and 14% of the database respectively. Approximately 0.4% (6 drugs) had origins other than those mentioned above, which consisted of those that encompass drugs of any origin, or drug monographs with no consensus on the origin. For example, "*So'out*" is a general term including any drug that is insufflated through the nose. Another example is "*Ghavand*", a drug of ambiguous origin. As Aghili has stated, some scholars have described it as an herbal oil, while others have believed it to be an oil of animal origin. Collectively, herbs constituted the most instances in all of types of *Mizaj*, whilst medicinal substances with an animal origin were mostly hot-wet, and minerals typically had a dry, cold-dry or balanced *Mizaj*.

Scientific Names

Scientific names from the SciResources 1-2 were stored in two separate fields. Collectively, a total of 1822 drug monographs possess scientific names from one or both resources. Fig. 2C provides information on the number of scientific names. As demonstrated, 1367 herbs, 301 drugs with animal origins, and 138 minerals had scientific names. There are 66 and 114 different families based on 'SciName1' and 'SciName2' respectively. Fig. 2D shows the most common plant families extracted from the two resources with the highest frequency seen in *Leguminosae*, *Compositae*, *Lamiaceae* and *Apiaceae* families, respectively. *Leguminosae*, commonly known as *Fabaceae* or pea family is widely distributed with approximately 730 genera and 19,400 species (39). This richness of this species is reflected in great morphological and chemical diversity, from which numerous compounds are derived (40). The outstanding feature of this family is the presence of roots tubercles or nodules which harbor nitrogen-fixing bacteria (*Rhizobium*). Pharmacological actions of these plants include analgesic,

anti-inflammatory, antiulcer, anticancer, antidiabetic, anti-inflammatory, antirheumatic, antimicrobial, antibacterial, and cytotoxic activities (41, 42). The Compositae (*Asteraceae*) is one of the largest and most diverse families of flowering plants. They are found in diverse habitats and contains over 40 economically important species including food (lettuce, Jerusalem artichoke), oil (sunflower, safflower), medicinal (chamomile) and ornamental (*Chrysanthemum*, *dahlia, zinnia, marigold*) crops. Many members of *Asteraceae* with pharmacological activity contain important phytochemical compounds such as polyphenols, flavonoids, and diterpenoids. and have antibacterial, antifungal, anti-inflammatory, insecticide, and antitumor capacities of *Asteraceae* species (43, 44). *Lamiaceae*, also known as the mint family is a large plant family of mostly shrubs and herbs contains about 236 genera and 6,900 to 7,200 species. *Lamiaceae* are distributed globally with particularly high diversity in the Mediterranean region. Medicinal properties of the *Lamiaceae* species are often attributed to their high content of volatile compounds. Many members of the family including Ajuga, Coleus, and Salvia are widely cultivated as ornamentals, while others such as sage (*Salvia*), thyme (*Thymus*), mint (*Mentha*), marjoram (*Origanum*), rosemary (*Rosmarinus*), lavender (*Lavandula*), and basil (*Ocimum*) are widely used as culinary herbs and spices (45). The *Apiaceae* or carrot family consists of plants with a characteristic umbrella-arranged fruit. Comprising of approximately 450 genera and 3700 species, this family of plants can be found worldwide (46). The noteworthy character of this family is the high content of volatile oil or oleoresin (47). Members of *Apiaceae* possess various compounds with many biological activities. Some of the main properties are antibacterial, hepatoprotective, vaso-relaxant, and antitumor activities, as well as the ability to induce apoptosis and inhibit cyclooxygenase (48, 49). Other medicinal herbs with high commercial value can be especially found in *Solanaceae, Boraginaceae, Cucurbitaceae, Euphorbiaceae* and *Papaveraceae* plant families.
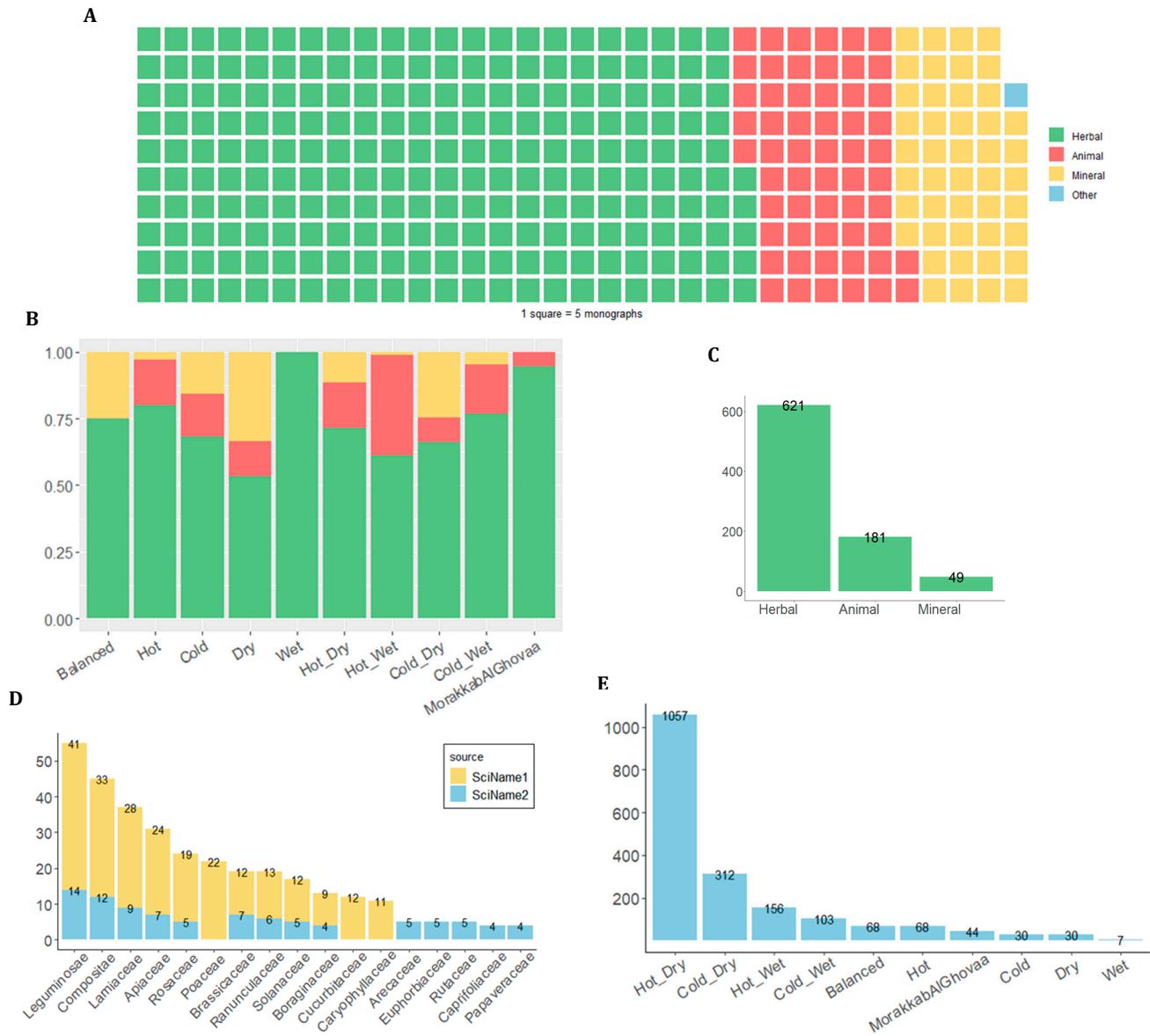
Figure 2: (A) Frequency of each origin (herbal, animal, mineral, other) in UNaProd database; (B) The share of each origin in the main ITM Mizaj types. (C) Frequency of scientific names extracted for drugs with herbal, animal, and mineral origins; (D) The most common plant families of UNaProd extracted from each of the scientific name resources. (E) Frequency of the main Mizaj types of drug monographs in UNaProd. The numbers on each bar represents the exact number of frequency.

Drug *Mizaj*

The *Mizaj* type and degree for every drug monograph with a *Mizaj* description in *Makhzan al-Advieh* were determined based on IrGO (Iranian traditional medicine General

Ontology), and thereafter reviewed manually by the ITM committee. Users can view the attributes of the specified *Mizaj* type and degree provided for each drug, as these fields were linked to the IrGO database. Moreover, in the instances where Aghili has quoted a different *Mizaj* from other physicians and pharmacologists, the name of the scholar together with the *Mizaj* type they have specified were provided.

Aghili has described a *Mizaj* for 1421 of the 1741 main drug monographs, including 384 views from other scholars. Considering the 2696 tuples of UNaProd, a total of 1976 tuples contain information regarding *Mizaj*. There are ten main *Mizaj* types in ITM, the frequency which was demonstrated in Fig. 2E. The most common *Mizaj* type for medicinal substances of *Makhzan al-Advieh* was the hot-dry *Mizaj* constituting 55.7% of the drug monographs with a determined *Mizaj*, followed by cold-dry (16.5%), hot-wet (8.1%), cold-wet (5.4%), balanced (4.8%), hot (3.6%), *Morakkab al-Ghovaa* - a *Mizaj* ascribed to drugs that can exhibit more than two qualities - (2.3%), dry (1.7%), cold (1.6%), and finally wet (0.4%) *Mizaj* types.

Regarding *Mizaj* degrees, ITM specifies an active and a passive quality for substances. Active qualities include hotness or coldness, while wetness and dryness constitute passive qualities. The potency of each quality is described in a range from first to fourth. The frequency of drug monographs with *Mizaj* types which have both qualities specified, i.e. hot-wet, hot-dry, cold-wet, and cold-dry; balanced; and *Morakkab al-Ghovaa Mizaj* types have been illustrated in Fig 3. Hot-dry drugs were mostly second or third degree in hotness (Fig. 3A), while for cold-dry and cold-wet drugs both active and passive qualities were in second degree (Fig. 3B and 3D), and the hotness and wetness in hot-wet drugs generally did not exceed second degree (Fig. 3C). The most common qualities in drugs with a balanced *Mizaj* inclined towards slight hotness and dryness (Fig. 3E). In drugs with *Morakkab al-Ghovaa Mizaj*, the passive quality of dryness was specified more prominently than wetness (Fig. 3F).
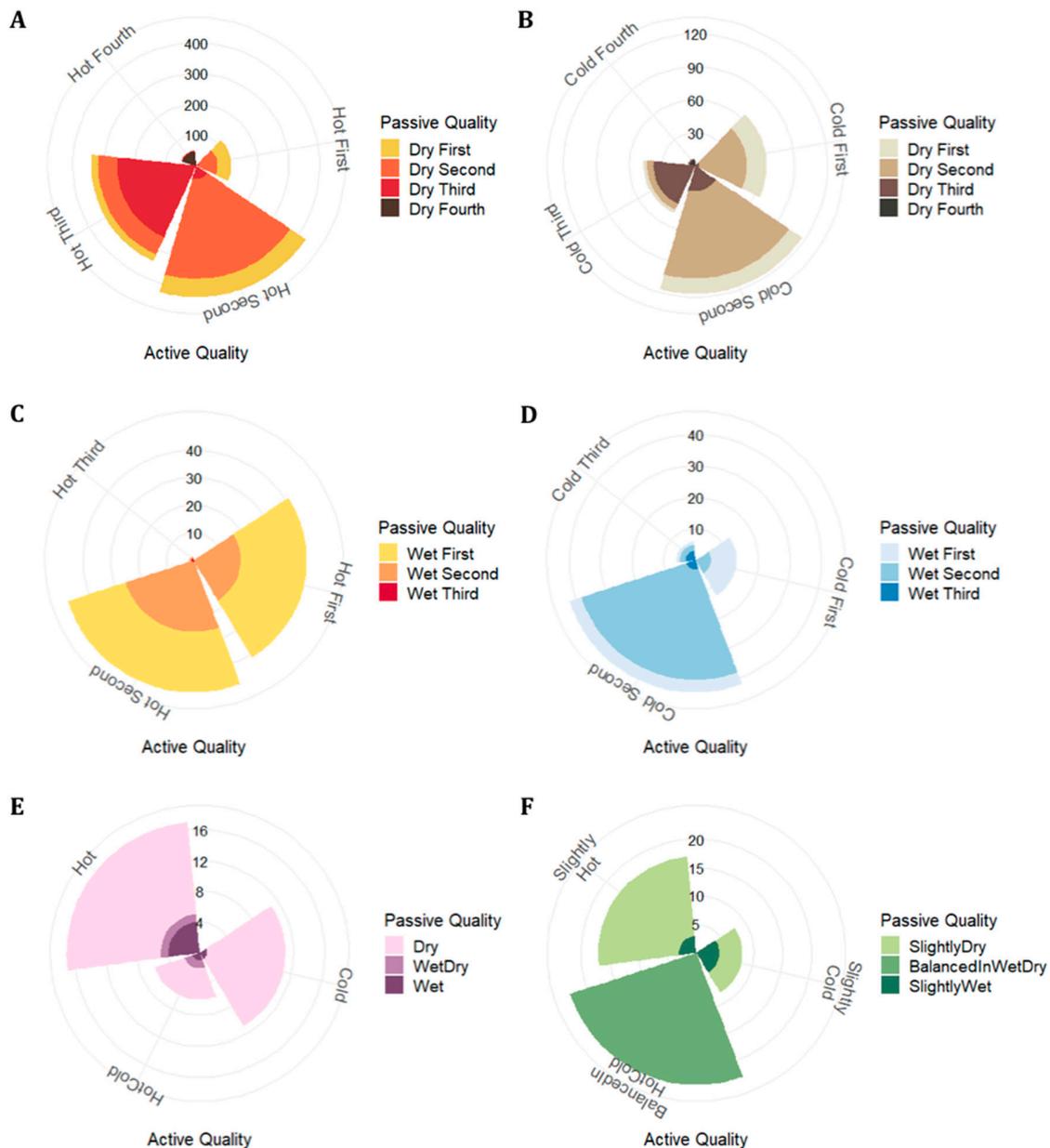
Figure 3: Mizaj degrees based on active (hotness/coldness) versus passive (wetness/dryness) qualities in Mizaj types of UNaProd drug monographs. In all plots the x axis is the active degree and the y axis is the count of drug monographs. (A) Hotness degree versus dryness degree in drugs with hot-dry Mizaj; (B) Coldness degree versus dryness degree in drugs with Cold-dry Mizaj; (C) Hotness degree versus wetness degree in drugs with hot-wet Mizaj; (D) Coldness degree versus wetness degree in drugs with Cold-wet Mizaj; (E) Active degree versus passive degree in drugs with balanced Mizaj; (F) Active degree versus passive degree in drugs with Morakkab al-Ghovaa Mizaj.

The *Mizaj* type and degree of UNaProd drug monographs have been also demonstrated in Fig. 4 in more detail. The active quality is in a range from coldness in maximum fourth degree to hotness in maximum fourth degree, while passive quality ranges from dryness in maximum fourth degree to wetness in maximum fourth degree. There are a number of

14

instances where Aghili has only described the *Mizaj* type but not the degree. For example, *Khatai* tea has a hot and dry *Mizaj* but neither the degree of hotness nor dryness have been specified. To demonstrate these drug monographs in the scatter plot, DNS factors have been added to each axis (See Table 1). Out of the total 1249 hotness degrees and 441 coldness degrees, 448 and 227 were unspecified respectively. Regarding passive qualities, there were 148 and 442 unspecified degrees out of the 386 and 1480 existing degrees for wetness and dryness degrees respectively.
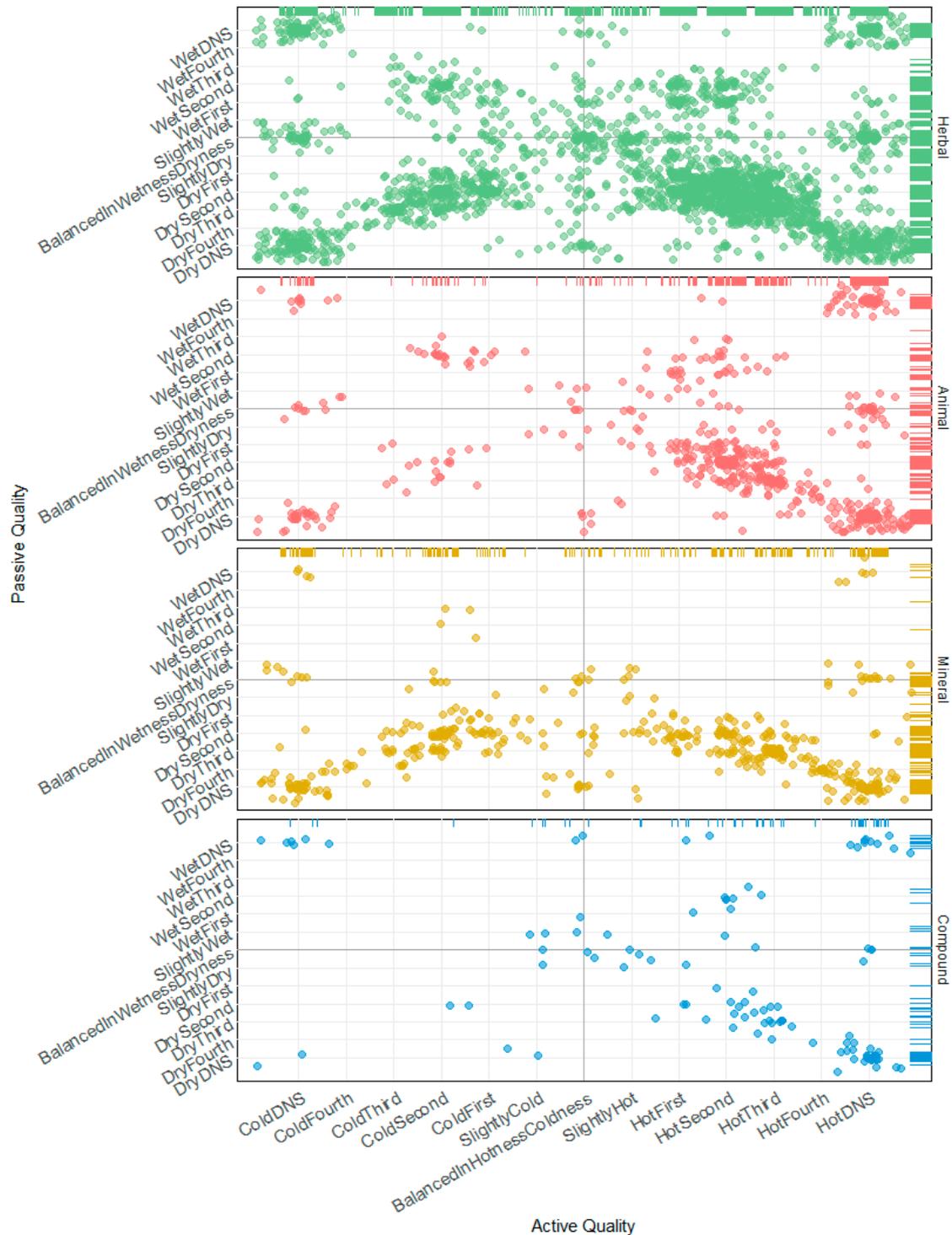
Figure 4: Mizaj jittered scatterplots of UNaProd drug monographs in terms of active versus passive quality faceted by origin. (A) herbal drug monographs; (B) animal drug monographs; (C) mineral drug monographs; (D) compound drug monographs. The two perpendicular lines in each plot indicate balance in either axis. The active quality consists of the main degrees of hotness and coldness, while passive quality includes the main degrees from fourth

degree in dryness to fourth degree in wetness. As some Mizaj descriptions do not contain the exact degree, a further factor - DNS (Degree Not Specified) has been defined for each quality.

User Interface

The home page of UNaProd (Fig. 5) database accommodates an introduction to UNaProd as well as the following features: (1) the 'database' link enables users to browse the drugs and their specifications via a table containing all fields. A search bar has been incorporated in this page that allows search in a specific field or all of the database; (2) A 'user manual' section describes the general structure of the database along with an explanation of each field and the keywords by which they have been extracted from the book; (3) 'Statistics': information regarding the integrity of the database and also statistics on the origin and Mizaj fields and the languages for which synonyms have been mentioned, can be accessed via a separate link in the database. The database is freely available to all users with no restriction at http://jafarilab.com/unaprod/.

Figure 5: UNaProd homepage including an introduction to UNaProd along with the following features: (1) the 'database' link; (2) a 'user manual'; and (3) 'Statistics'.

18

Conclusion and Future Perspective

As a traditional medical system with a prolific history, ITM contains knowledge that can be better utilized if promoted using modern systems medicine approaches. Meanwhile, the rationale behind physiologic or pathologic processes can be nowadays made possible by the analysis of complex networks (17). Due to the holistic nature of ITM, deducing justifications for the concepts of the personalized approach to maintain health and treating diseases requires establishing links to modern medicine. Thus, standardization and organization of ITM concepts and drugs shall undoubtedly facilitate the validation of its theories, as has been made possible with other traditional medical systems (31).

An outset to the standardization and systematization of ITM is creating an open knowledge base which can subsequently be linked with modern biological databases to fulfill the increasing demands of researchers in achieving a deeper understanding of the concepts and mechanisms of ITM. We have endeavored to take a step towards fulfilling this goal by creating a database of *materia medica*, UNaProd.

The knowledge and experience provided by traditional systems of medicine can help reduce the expenses, time consumed and also toxicity occurrences in developing new drugs from natural sources (50, 51). Thus, considering all the challenges and financial issues in the pharmaceutical industry, an effective utilization of this database may facilitate the drug discovery process from ethnomedical data. This process starts with selecting a drug from the *materia medica* (52), which may be accelerated by a database considering the large number of available information based on ITM.

Since IrGO ontologies include English translations and definitions for ITM concepts, we intend to link our database to ontologies of IrGO as they develop to make the database usable for non-Persian speakers. This will also be of aid in creating a systems pharmacology platform of ITM *materia medica* that encompasses the relationships between the drugs, their targets and diseases; and constructing a network of the information provided in ITM literature. We will endeavor to expand the database and include other resources of ITM *materia medica* and also provide links to phytochemical and other traditional medicine databases.

Supplementary file legends

Supplementary file 1: UNaProd database diagram. The 'Drug' table contains the primary fields extracted from the text of Makhzan al-Advieh i.e. 'Identity', 'Mizaj', 'ActionsandMedicinalUses', 'Dosage', 'AdverseEffects', 'Refinements', 'Substitute' and 'Expression', along with IDs for drugs. It also includes 'Origin' extracted from the identity field; and 'CommonName', 'SciName1', and 'SciName2' extracted from other resources. The table 'DrugMizaj' contains the Mizaj type and degree for the drug monographs extracted from IrGO database. The 'Synonyms' table includes the synonyms and translations of the drugs extracted from the identity and Expression fields. There are separate tables for origin types, languages, and scholar names mentioned in the book.

References

1.   Longo DL. Harrison's principles of internal medicine. New York: McGraw-Hill; 2012.

2.   World Health Organization. WHO traditional medicine strategy: 2014-2023. World Health Organization, http://wwwwhoint/iris/handle/10665/92455 2013

3.   Shirbeigi L, Zarei A, Naghizadeh A, Alizadeh Vaghasloo M. The Concept of Temperaments in Traditional Persian Medicine. Trad Integr Med. 2017;2(3):143-56.

4.   Zarshenas MM, Mehdizadeh A, Zargaran A, Mohagheghzadeh A. Rhazes (865–925 AD). Journal of Neurology. 2012;259(5):1001-2.

5.   Zargaran A, Mehdizadeh A, Zarshenas MM, Mohagheghzadeh A. Avicenna (980–1037 AD). Journal of Neurology. 2012;259(2):389-90.

6.   Moeini R, Gorji N. Persian Medicine in the World of Research; Review of Articles on Iranian Traditional Medicine. Iranian journal of medical sciences. 2016;41(3 Suppl):S7.

7.   Jing T, Tero A. Network Pharmacology Strategies Toward Multi-Target Anticancer Therapies: From Computational Models to Experimental Design Principles. Current Pharmaceutical Design. 2014;20(1):23-36.

8.   Jafari M and Karimi M. Systems biology recalls to us some aspects of traditional medicine [version 1; not peer reviewed]. F1000Research 2018, 7:1080 (slides) (https://doi.org/10.7490/f1000research.1115824.1).

9.   Bose B. Systems biology: a biologist's viewpoint. Progress in biophysics and molecular biology. 2013;113(3):358-68.

10.   Chen L. Systems biology with omics data. Methods (San Diego, Calif). 2014;67(3):267-8.

11.   Chuang HY, Hofree M, Ideker T. A decade of systems biology. Annual review of cell and developmental biology. 2010;26:721-44.

12.   Kirschner M. Systems Medicine: Sketching the Landscape. Methods in molecular biology (Clifton, NJ). 2016;1386:3-15.

13.   Kitano H. Systems biology: a brief overview. Science (New York, NY). 2002;295(5560):1662-4.

14.   Lucas M, Laplaze L, Bennett MJ. Plant systems biology: network matters. Plant, cell & environment. 2011;34(4):535-53.

15.   Prathipati P, Mizuguchi K. Systems Biology Approaches to a Rational Drug Discovery Paradigm. Current topics in medicinal chemistry. 2016;16(9):1009-25.

16.   Saqi M, Pellet J, Roznovat I, Mazein A, Ballereau S, De Meulder B, et al. Systems Medicine: The Future of Medical Genomics, Healthcare, and Wellness. Methods in molecular biology (Clifton, NJ). 2016;1386:43-60.

17.   Schneider HC, Klabunde T. Understanding drugs and diseases by systems biology? Bioorganic & medicinal chemistry letters. 2013;23(5):1168-76.

18.   Schneider MV. Defining systems biology: a brief overview of the term and field. Methods in molecular biology (Clifton, NJ). 2013;1021:1-11.

19.   Karbalaei R, Piran M, Rezaei-Tavirani M, Asadzadeh-Aghdaei H, Heidari MH. A systems biology analysis protein-protein interaction of NASH and IBD based on comprehensive gene information. Gastroenterology and hepatology from bed to bench. 2017;10(3):194-201.

20.   Karbalaei R, Allahyari M, Rezaei-Tavirani M, Asadzadeh-Aghdaei H, Zali MR. Protein-protein interaction analysis of Alzheimer`s disease and NAFLD based on systems biology methods unhide common ancestor pathways. Gastroenterology and hepatology from bed to bench. 2018;11(1):27-33.

21.   Jafari M, Rezadoost H, Karimi M, Mirzaie M, Rezaie-Tavirani M, Khodabandeh M, et al. Proteomics and traditional medicine: new aspect in explanation of temperaments. Forsch Komplementmed. 2014;21(4):250-3.

22.   Rezadoost H, Karimi M, Jafari M. Proteomics of hot-wet and cold-dry temperaments proposed in Iranian traditional medicine: a Network-based Study. Scientific Reports. 2016;6:30133.

23.   Suo T, Wang H, Li Z. Application of proteomics in research on traditional Chinese medicine. Expert review of proteomics. 2016;13(9):873-81.

24.   Shaneyfelt ME, Burke AD, Graff JW, Jutila MA, Hardy ME. Natural products that reduce rotavirus infectivity identified by a cell-based moderate-throughput screening assay. Virology Journal. 2006;3(1):68.

25.   Cook D, Lee ST, Taylor CM, Bassüner B, Riet-Correa F, Pfister JA, et al. Detection of toxic monofluoroacetate in Palicourea species. Toxicon. 2014;80:9-16.

26.   Obbo CJD, Makanga B, Mulholland DA, Coombes PH, Brun R. Antiprotozoal activity of Khaya anthotheca, (Welv.) C.D.C. a plant used by chimpanzees for self-medication. Journal of Ethnopharmacology. 2013;147(1):220-3.

27.   Shaherin B, Minghua C, Stephani JYM, Sun C. Expediting the Design, Discovery and Development of Anticancer Drugs using Computational Approaches. Current Medicinal Chemistry. 2017;24(42):4753-78.

28.   Fakhrudin N, Waltenberger B, Cabaravdic M, Atanasov AG, Malainer C, Schachner D, et al. Identification of plumericin as a potent new inhibitor of the NF-κB pathway with anti-inflammatory activity in vitro and in vivo. British Journal of Pharmacology. 2014;171(7):1676-86.

29.   Pan S-Y, Zhou S-F, Gao S-H, Yu Z-L, Zhang S-F, Tang M-K, et al. New Perspectives on How to Discover Drugs from Herbal Medicines: CAM's Outstanding Contribution to Modern Therapeutics. Evidence-Based Complementary and Alternative Medicine. 2013;2013:25.

30.   Gyllenhaal C, Kadushin MR, Southavong B, Sydara K, Bouamanivong S, Xaiveu M, et al. Ethnobotanical approach versus random approach in the search for new bioactive compounds: support of a hypothesis. Pharmaceutical biology. 2012;50(1):30-41.

31.   Huang L, Xie D, Yu Y, Liu H, Shi Y, Shi T, et al. TCMID 2.0: a comprehensive resource for TCM. Nucleic acids research. 2018;46(D1):D1117-D20.

32.   Kan WK, Siu YT, But PPH. TCMD: An On-Line Database of Traditional Chinese Medicine. Proceedings of the AMIA Annual Fall Symposium. 1996:821-.

33.   Zeng X, Zhang P, Wang Y, Qin C, Chen S, He W, et al. CMAUP: a database of collective molecular activities of useful plants. Nucleic acids research. 2019;47(D1):D1118-D27.

34.   Petrovska BB. Historical review of medicinal plants' usage. Pharmacognosy reviews. 2012;6(11):1-5.

35.   Aghili M. Makhzan-al-Advieh. Tehran: Tehran University of Medical Sciences: Rah-e Kamal; 2009.

36.   Ghahreman A, Okhovat A. Matching the Old Medicinal Plant Names with Scientific Terminology [In Persian]. Tehran, Iran: University of Tehran Publications; 2004.

37.   Naghizadeh A, Akbari S, Hamzeian D, Rezaeizadeh H, Alizadeh Vaghasloo M, Mirzaie M, et al. IrGO 1.0: Iranian traditional medicine General Ontology. F1000Research. 2018;7:1319-.

38.   Khallash M, Imany M. Hazm: Python library for digesting Persian text: Github; 2018 [Available from: https://github.com/sobhe/hazm.

39.   Judd WS, Campbell CS, Kellogg EA, Stevens PF, Donoghue MJ. Plant systematics: Sinauer Sunderland, Massachusetts, USA; 2002.

40.   Waterman PG. Costs and benefits of secondary metabolites to the Leguminosae. Advances in Legume Systematics. 1994;5:129-49.

41.   Molares S, Ladio A. The usefulness of edible and medicinal Fabaceae in Argentine and Chilean Patagonia: environmental availability and other sources of supply. Evidence-Based Complementary and Alternative Medicine. 2012;2012.

42.   Gao T, Yao H, Song J, Liu C, Zhu Y, Ma X, et al. Identification of medicinal plants in the family Fabaceae using a potential DNA barcode ITS2. Journal of ethnopharmacology. 2010;130(1):116-21.

43.   Singh B, Sahu P, Sharma M. Anti-inflammatory and antimicrobial activities of triterpenoids from Strobilanthes callosus Nees. Phytomedicine. 2002;9(4):355-9.

44.   Koc S, Isgor BS, Isgor YG, Shomali Moghaddam N, Yildirim O. The potential medicinal value of plants from Asteraceae family with antioxidant defense enzymes as biological targets. Pharmaceutical biology. 2015;53(5):746-51.

45.   Khoury M, Stien D, Eparvier V, Ouaini N, El Beyrouthy M. Report on the medicinal use of eleven Lamiaceae species in Lebanon and rationalization of their antimicrobial potential by examination of the chemical composition and antimicrobial activity of their essential oils. Evidence-Based Complementary and Alternative Medicine. 2016;2016.

46.   Pimenov MG, Leonov MVe. The genera of the Umbelliferae: a nomenclator: Royal Botanic Gardens, Kew; 1993.

47.   Singh V. Taxonomy of angiosperms: Rastogi publications; 1981.

48.   Amiri MS, Joharchi MR. Ethnobotanical knowledge of Apiaceae family in Iran: A review. Avicenna journal of phytomedicine. 2016;6(6):621.

49.   Pae HO, Oh H, Yun YG, Oh GS, Jang SI, Hwang KM, et al. Imperatorin, a furanocoumarin from Angelica dahurica (Umbelliferae), induces cytochrome c-dependent apoptosis in human promyelocytic leukaemia, HL-60 cells. Pharmacology & toxicology. 2002;91(1):40-8.

50.   Pan S-Y, Zhou S-F, Gao S-H, Yu Z-L, Zhang S-F, Tang M-K, et al. New Perspectives on How to Discover Drugs from Herbal Medicines: CAM's Outstanding Contribution to Modern Therapeutics. Evidence-based complementary and alternative medicine : eCAM. 2013;2013:627375-.

51.   Kibble M, Saarinen N, Tang J, Wennerberg K, Mäkelä S, Aittokallio T. Network pharmacology applications to map the unexplored target space and therapeutic potential of natural products. Natural Product Reports. 2015;32(8):1249-66.

52.   Fabricant DS, Farnsworth NR. The value of plants used in traditional medicine for drug discovery. Environmental health perspectives. 2001;109 Suppl 1:69-75.