

植物资源与化学数据库系统(DPRC) 的功能与应用

贺震旦¹ 施庆华² 赵玺² 张颖君¹ 杨崇仁^{1*}

(¹中国科学院昆明植物研究所, 昆明 650204, ²云南省机械研究设计院)

摘要 由中国科学院昆明植物研究所和云南省机械研究设计院研制的植物资源与化学数据库系统(DPRC)是一个将植物资源、植物化学成分及化学结构的多种信息、数据和图象融为一体的、开放式多功能软件系统,包括植物资源检索系统(KSPR)、植物化学成分检索系统(KSPC)及天然产物化学结构检索系统(KSCS)等三个子系统。各子系统都是开放型的,能不断输入、修改和补充存储数据,既可独立运行,又能相互联系形成一个整体。

关键词 植物资源;植物化学成分;天然产物化学结构;数据库;模糊检索

The function and application of the Database of Plant Resources and Chemistry He Zheng-Dan¹, Shi Qing-Hua², Zhao Xi², Zhang Ying-Jun¹ and Yang Chong-Ren^{1*} (¹Kunming Institute of Botany, Chinese Academy of Sciences, Kunming 650204, ²Yunnan Institute of Machine), *J. Plant Resour. & Environ.* 1995, 4(3): 49~55

The function and application of a new developed computer software system—Database of Plant Resources and Chemistry (DPRC) are presented. It combined the informations of plant resources, the chemical and physical and spectrum data of compounds from plants and the structures of these constituents to one system which is an opening polyfunctional software system. This system is composed of Key System of Plant Resources (KSPR), Key System of Plant Chemical Constituents (KSPC) and Key System of Chemical Structure of Natural Products (KSCS). Each daughter system is an opening system which could be inputted, revised and saved data constantly. The three daughter systems could be used not only individually but also combinable and form an integrated system.

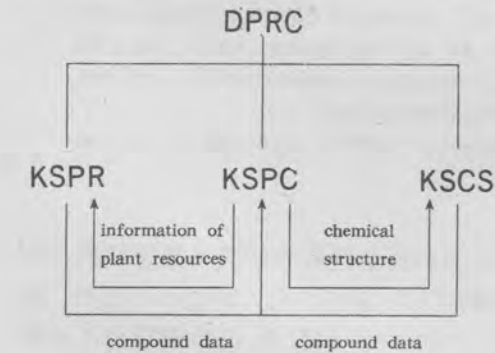
Key words plant resources; plant chemical constituents; chemical structure of natural products; database; vaguely searching data

植物资源是各种天然资源中内容最丰富,与国计民生和社会经济生活联系最密切的一类可再生的资源。为充分评价及合理开发、利用和保护植物资源,除进行深入调查和研究外,还需对浩如烟海的植物学和植物化学文献资料进行整理、归纳、分析与综合。这样艰巨的工作没有一大批治学严谨且有丰富经验的专家的长期辛勤工作是不可能完成的,而这还需要庞大的组织与协调。

随着电子计算机技术的发展,微机在各学科领域的应用日益普及,利用电子计算机存储和处理大量植物资源和天然产物化学信息与数据,提供检索服务已成为可能。迄今,国内外在生物学、农林和医学等方面开发成功的数据库已近200个。例如:在植物资源方面,中国科

学院计算机中心科学数据库研制了“中国药用植物数据库系统”(CMP)。香港中文大学建立了中药数据库。在化合物方面,英国 Chapman & Hall 公司推出了“天然化合物辞典”光盘(The Dictionary of Natural Products CD-ROM)。上海药物研究所建立了“天然产物数据系统”。广州化学研究所建立了“化合物 SANTELER 标准¹³C NMR 光谱编号检索系统”,此外,国内还建立有“红外光谱数据库”,“化合物命名与结构数据库”以及“有机化合物谱图数据库”等。至于化学结构图,通用的有 IBM 系列的 CAD 和 SIGMA 公司的 WIWP 图形软件以及 MACINTOSH 系列的 CHEMDRAWH CHEM. 3D 等,均具有化学结构制图的各种功能并已商品化。然而这些数据库大多都不能任意删减。迄今为止,尚未见到一个将植物资源、化合物及化学结构的多种信息、数据和图象融为一体、开放式的多功能软件系统。为了建成一个具有足够丰富的信息量和多项选择检索功能,既可以成为植物资源和植物化学工作者日常工作的助手,又是一个可以随时提供服务的综合知识库,在植物资源的分析评价、化学成分结构解析以及植物学和药用植物学的一些研究领域均能起到辅助作用的综合性多功能软件系统。我们研制了“植物资源与化学数据库系统”(Database of Plant Resources and Chemistry)(DPRC)微机软件系统。该系统包括三个子系统(图1),它们的功能与应用分述如下:

1. 植物资源检索系统(Key System of Plant Resources)(KSPR) 在植物资源和植物学一些分支领域以及药用植物学的研究中,常常需要收集某些地区、或具有某种用途或特性、或含有某些化合物的某类植物资源的有关信息,进行分析整理并列表统计。这些大量的信息分散在许多文献中,即使用抄卡片的方式也难于分析和整理。KSPR 系统为这项工作提供了便利的工具。该系统以 FOXBASE 语言为基础,经与 UC DOS, CCDOS 2.13 等汉字软件系统联用,可同时输入有关植物资源的多项中英文信息,并具有增删、修改、检索、打印和拷贝等功能(图2)。



KSPR: Key System of Plant Resources
KSPC: Key System of Plant Chemical Constituents
KSCS: Key System of Chemical Structure of Natural Products

图1 DPRC 系统的三个子系统

Fig 1 The three daughter systems of DPRC

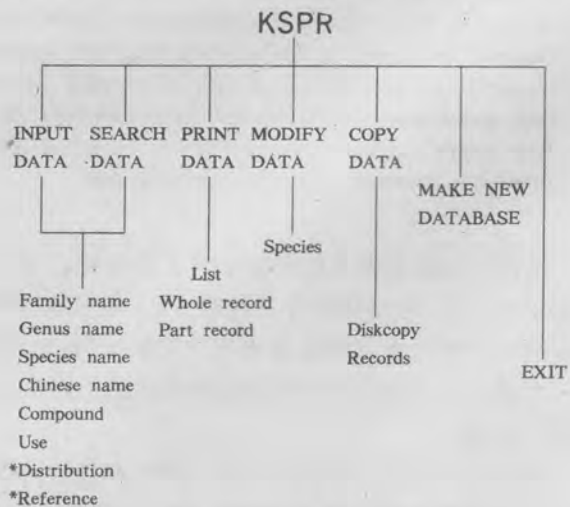


图2 KSPR 子系统的主要功能

Fig 2 The main function of KSPR

KSPR 系统收录的植物资源信息包括植物的科名、属名、种名、中文名、化学成分、用途、分

布以及文献等。有时,还可根据需要从中文名、化学成分和用途等项目中输入植物习性、分布区类型、利用部位、采集季节、蕴藏量等信息。其中植物科名、属名、种名、中文名、化学成分和用途等项目可进行检索。除单项检索外,植物的科名、属名和种名中的任一项还可分别与中文名、化学成分、用途和分布中的任一项组合检索。检索结果可显示清单及记录内容并进行打印。只要存储有大量的信息,KSPR系统即可进行植物资源的多指标综合分析,提供量化的分析结果。成为植物资源的评价与开发利用以及植物区系地理、植物化学分类学和药用植物学等学科领域研究中的得力工具和助手。为充分发挥KSPR系统的作用,在使用过程中应注意的关键问题是:

(1) 信息的输入 我国有种子植物三万余种,还有大量的孢子植物,有关植物资源的信息散见于各种古今文献和数百种杂志资料中。要一时输入如此繁杂的植物资源信息,工作量是十分巨大的。而KSPR系统检索和分析的可靠性与实用性又与输入信息的量成正比。因此,每一用户宜根据自己工作的需要,从某几个植物类群(某几个科)或具有某些特殊性状(如用途、分布等)植物资源入手建立KSPR信息库,再逐步扩大收录的范围。

(2) 数据模式化 KSPR系统与其他微机软件系统一样,对于信息的分析和检索是通过模式识别功能进行的。因此,信息输入格式的模式化是十分重要的。不经过模式化输入的信息如同一堆乱草,KSPR系统无法识别、分析和检索。然而,分散在各种文献中的植物资源信息又是十分不规范的。这就要求用户在输录过程中通过判断进行模式加工,使输入的信息具有统一的格式。KSPR系统对一些信息的输入格式已作了规定。例如,植物的科名、属名和种名均要求按拉丁学名输入,其中,种名一项要求按双名法输入完整的属名和种名,定名人名称的表示方式因无法规范(如Franch.和Fr.等),不进行检索,可根据需要决定输入与否。对另一些字段,则需用户自行设置约定俗成的格式。例如:植物的中文名中少数民族的植物名称,以统一用汉语拼音字母为宜,这比通常的汉字音译更易格式化;化合物名若统一用英文输入,必要时在后面用括号记入中文名,则将便于输录和检索;用途一项的模式化比较复杂,宜根据具体的情况进行,如可按植物资源的大类用途格式为香料、药物、油料、纤维、淀粉、丹宁、饲料……等,也可在药用植物中按用途格式为抗癌、抗炎、抗菌、镇静、止痛……等,也可按中医理论的性味、功能、归经和主治等格式,也可按使用的民族格式为傣族药、彝族药、白族药、蒙古族药……等。总之,应根据需要和实际情况设置不同层次、不同内容的输入格式。模式化越科学,KSPR系统的功能越可能充分发挥。该系统检索功能的示范如图3。

2. 植物化学成分检索系统(Key System of Plant Chemical Constituents) (KSPC) 在天然有机化学研究中,鉴定一个化合物的分子结构需要参考大量的物理化学数据和光谱数据,对于一个不常从事某类化合物结构研究的人来说,解析一张谱图仍然是件很困难的工作。即使是熟练的工作者,要核对各项数据也十分繁琐而耗时。由KSPC系统建立的化学信息库为天然化合物的结构解析提供了有力的帮助。该系统以FOXBASE语言为基础,可以英文方式输入化合物的各项物理化学常数和光谱数据以及有关信息,并具有增删、修改、检索、打印和拷贝等项功能(图4)。

KSPC系统的一个特点是输入内容丰富,包括化合物的名称、化学名、分子式、分子量、熔点、比旋光度、紫外光谱、红外光谱、 ^1H 核磁共振谱、 ^{13}C 核磁共振谱、质谱(包括EI-MS, FD-MS

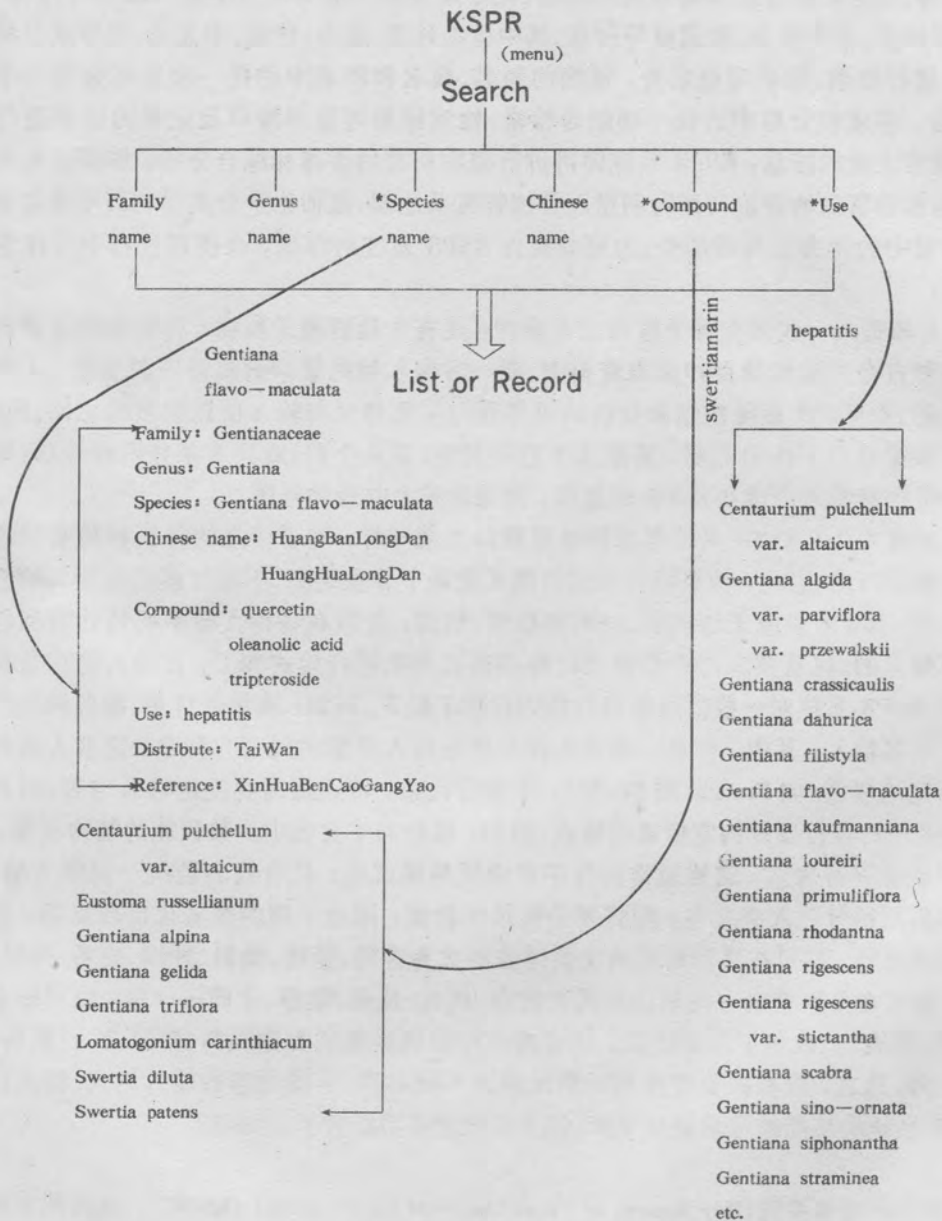


图3 KSPR 子系统的检索实例

Fig 3 The searching example of KSPR

和 FAB-MS)、生理活性、植物来源以及文献等大量的数据和信息,而且所有的物理化学常数和光谱数据均可进行检索。同时, KSPC 系统还可对各项数据进行模糊检索,只要设置一定的误差范围,就能通过系统分析检出符合条件的类似化合物,从而具有一定程度的人工智能化特征。这不仅有利于已知化合物的鉴定,对新化合物的结构解析也将起到辅助作用。显然,一旦收录了大量信息, KSPC 系统不仅是一个天然化合物的数据库,而且将成为植物化学工作者日常的助手。和 KSPR 系统一样,信息的输入量和数据的模式化也仍然是 KSPC 系统发挥

其特长和作用的关键问题。

为便于检索和系统分析, KSPC 对各项数据的输入均建立了一定的模式。例如, 化合物名称用英文小写字母, 分子式按 CHON 和其他杂原子顺序排列, 比旋光值正值不写“+”号, 负值则在数据前加“-”号; 红外光谱、紫外光谱及质谱数据均需按大小顺序排列等。各数据之后均可用括号说明必要的信息, 如紫外光谱的克分子消光值, 核磁共振谱化学位移的峰形、偶合常数及归属、文献等, 括号内容不进行检索。通过严格的输入格式与灵活的多种信息收录相结合, KSPC 系统提供的信息就相当丰富了。

KSPC 系统对于核磁共振谱化学位移的模糊检索使该系统向智能化迈出了一大步。只要将图谱数据由大到小输入, 并设置一定误差值, 系统即可进行分析, 由于仪器或溶剂产生的误差或甚至因叠峰而造成的信号个数误差均能通过模糊检索而克服。

该系统的检索功能, 特别是模糊检索功能示范如图5及图6。

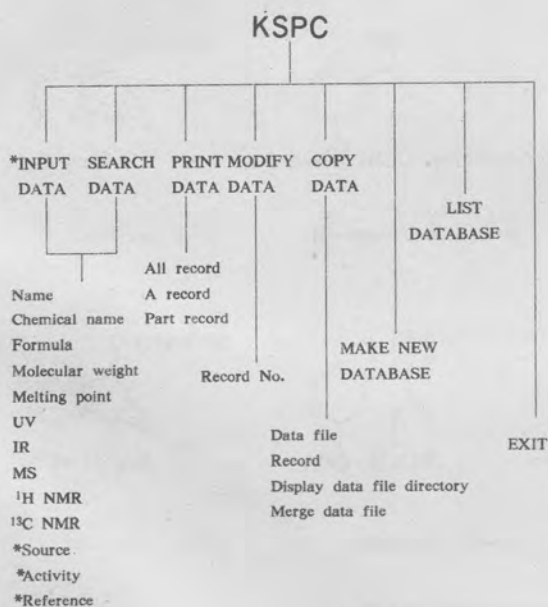


图4 KSPC子系统的主要功能
Fig 4 The main function of KSPC

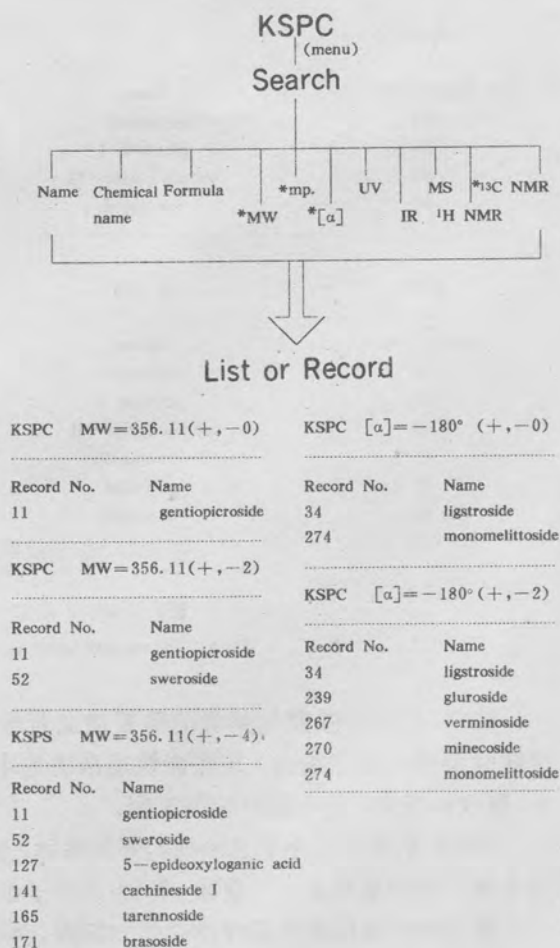


图5 KSPC子系统的模糊检索实例

Fig 5 The vaguely searching example of KSPC

3. 化学结构检索系统(Key System of Chemical Structure of Natural Products)(KSCS) 该子系统是以 PASCAL 语言为基础, 具有化学结构图形生成、显示和绘制的功能(图7)。可与 KSPC 连用, 成为 KSPC 系统的补充。

^{13}C NMR(ppm): 168.8 161.7 147.3 142.4 136.8
131.3 117.0 114.5 103.0 99.9
95.2 81.4 78.7 77.8 74.9
71.8 66.9 63.0 61.3 60.3
43.3 36.8

KSPC +, -0

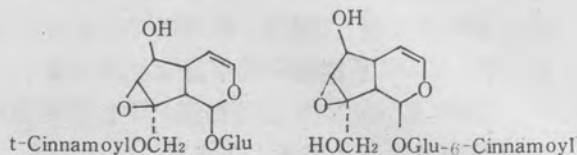
Record No.	Name
42	specioside

KSPC +, -5

Record No.	Name
35	globularin
37	picroside I
39	scutellarioside-II
42	specioside

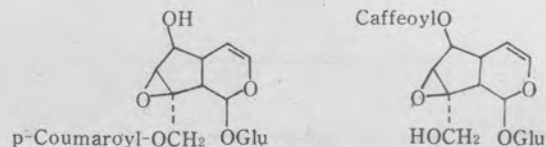
KSPC +, -9

Record No.	Name
35	globularin
37	picroside I
39	scutellarioside-II
41	verminoside
42	specioside
44	minecoside



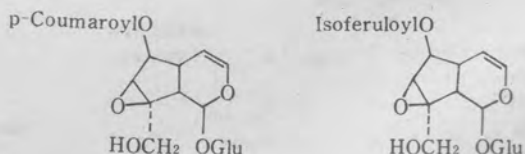
35 globularin

37 picroside I



39 scutellarioside-II

41 verminoside



42 specioside

44 minecoside

图6 KSPC 子系统 ^{13}C NMR 模糊检索实例

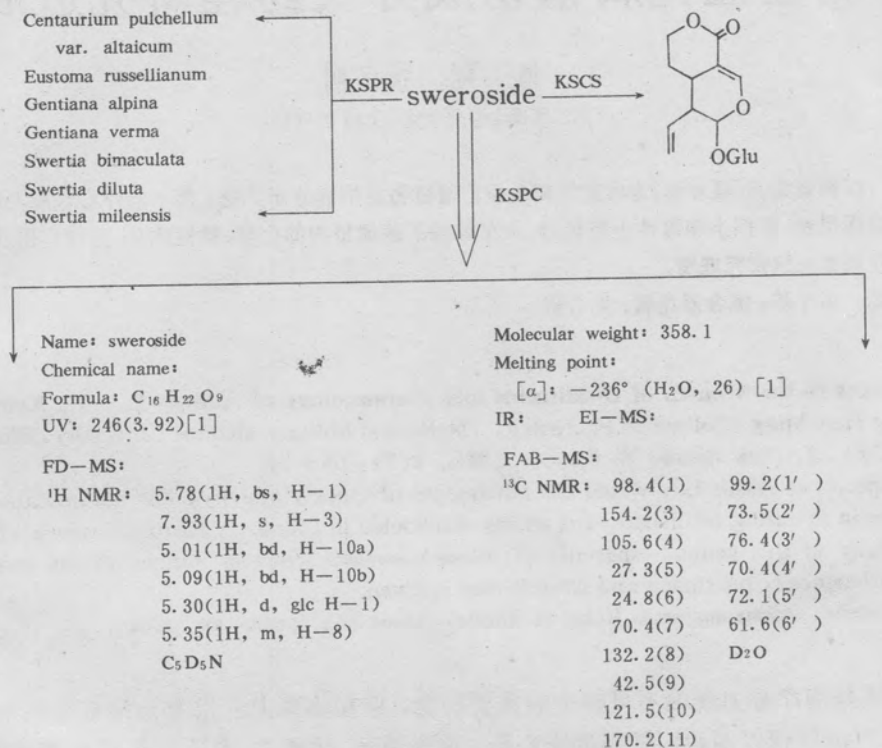
Fig 6 The vaguely searching example of ^{13}C NMR of KSPR

KSCS 系统可按照使用者的需要建立若干结构单元(母核)模块和官能团基团单元模块, 供随时调用和任意组合, 从而能快速准确地生成结构图。KSCS 建立的结构图可任意放大、缩小、删改和存储, 也可随时绘制成图。

DPRC 系统的三个子系统都是开放型的, 能不断输入、修改和补充存储数据, 既可独立运行又能互相联系形成一个整体。图7为三个子系统相互联系形成一个整体的实例。

由 DPRC 微机软件系统建立的知识库, 不仅可为植物资源和植物化学研究文献资料和光谱数据的收集整理提供迅速、简便的手段, 其多项检索和化学结构制图功能也为植物资源的综合分析和评价提供了有力的工具, 使化合物结构的鉴定迈出智能化的一步, 大大提高研究工作效率和决策水平。它不仅是一个可供检索的实用性强的数据库, 而且是从事植物资源和

植物化学研究的专家日常工作的得力助手。



Source:

1. Swertia (Gentianaceae)[1]
2. Eustoma russellianum [6]
3. Anthocleista (Loganiaceae)

4. Stereochemistry [4]
5. Mentzelia (Lossaceae) [5]
6. Dipsacus sylvestris (Dipsacaceae) [7]

Reference:

1. Tetrahedron Lett., No. 43, 5229(1966);
2. Phytochem. 18, 273(1979)
3. "Cyclopentanoid Terpene Derivatives", Marcel Dekker Inc, New York N. Y. 1969, Chapter 1
4. Chem. Pharm. Bull 18(9), 1889(1970)
5. L. J. El-Naggar, Ph. D. thesis, Ohio State University, 1980

Activity:

图7 KSCS 子系统的主要功能
Fig 7 The main function of KSCS

(责任编辑: 许定发)