

中药有效成分三维结构数据库的开发和研究

侯廷军 乔学斌 徐筱杰*

(北京大学化学与分子工程学院 北大养生堂天然药物研究室 北京 100871)

摘要 介绍了北京大学中药有效成分三维结构数据库软件系统的结构、功能及开发步骤. 该数据库系统不仅仅提供 6500 个中草药有效成分的二维和三维结构以及其它各类相关信息. 同时拥有功能强大的数据库查询、维护及分子表达系统. 在该系统中, 用户可以交互式地实现多种分子特征的查询以及二维子结构的查询. 查询得到的分子可以直接在北京大学药物设计系统(PKUDDS)中进行三维结构的显示和分析. 该数据库系统和我们科研组开发的北京大学药物设计系统以及中草药信息系统构成了完整的基于中药的药物设计系统. 该系统已经用于 NS3 - NS4A 蛋白酶抑制剂以及其它体系的研究并取得了很好的结果.

关键词 中药, 有效成分, 数据库, 三维结构, 子结构匹配, 北京大学药物设计系统

Research and Development of 3D Molecular Structure Database of Traditional Chinese Drugs

HOU Ting - Jun QIAO Xue - Bin XU Xiao - Jie*

(Beida Yangshengtang laboratory for Natural Products,
College of Chemistry and Molecular Engineering, Peking University, Beijing, 100871)

Abstract In this paper, the structure, function and development of the 3D Molecular Structure Database of Traditional Chinese Drugs (3D - MSDT) are introduced. The database system not only includes the 2D and 3D structures and other concerned information for about 6500 effective elements in Traditional Chinese Drugs, but also bears powerful database maintenance and molecular visualization functions. In this database system, the users can carry out the queries of many types of molecular properties and 2D substructures matching. The obtained molecules can be transferred in to the Peking University Drug Design System (PKUDDS) for molecular visualization and analyses. The 3D - MSDT system, combined with the Traditional Chinese Medicine Drug Information System and the Peking University Drug Design System developed in our group constitute a integrated drug design system based on Traditional Chinese Drugs. This system has been applied to discover new potential inhibitors of NS3 - NS4A protease, and the results are very encouraging.

Keywords Traditional Chinese Drug, effective elements, database, 3D structure, substructure matching, Peking University Drug Design System

中国中药资源丰富, 品种繁多, 应用历史悠久, 为人类社会的发展和进步作出了重要的贡献. 目前
国际市场上每年药用植物及相关产品的交易额达数百亿美元, 而且还呈逐年上升的趋势. 但作为中药发

* E-mail: xiaojxu@chem.pku.edu.cn

收稿日期: 2001 - 05 - 04, 修回日期: 2001 - 08 - 07, 定稿日期: 2001 - 08 - 15, 国家自然科学基金(29873003)资助项目

(Received May 24, 2001 Revised August 7, 2001 Accepted August 15, 2001)

源地的中国,其交易量还不到世界市场的5%,而其他一些国家,如日本、韩国、德国等,无论是在中药的开发、研制,还是市场占有率上都远远超过中国,这种情况与中国所具有的中药大国地位是极不相称的。造成这种局面的原因是多方面的,首先是中西方文化上的差异,中医药在中国经过了几千年的发展,形成了一套独特的理论体系,这些理论体系基本上是经验性的,很难得到国际广泛的认可和接受,其次是科技的差异,在客观上,中药从原材料到产品都还缺乏明确、严格的质量标准和规范,还缺乏采用现代科学技术手段来充分说明中药作用的有效成份、作用机理、药效特性、生物代谢等丰富的科学内涵。

目前中药开发研究有两条途径,一条途径是从单一植物提取一种有效成分或提取物开发成新药,前者如从青蒿中提取的青蒿素,再经过结构修饰,成为一系列治疟疾的良药,后者如从银杏叶中所得提取物,由多种银杏黄酮甙和银杏内酯组成,另一条途径是中药复方制剂的开发研究,对传统复方进行简化以及设计组合新的复方,但无论哪条途径,开发中药总的趋势是坚持化学成分研究和药效药理研究相结合的原则,从中草药的有效化学成分入手,结合药理研究,是实现中药现代化的一条必由之路。

中医药各类资源纷繁复杂,中药信息的管理和利用是中药现代化的一项重要内容,国际上各类结构数据库已成为新药开发的必备工具,各大制药公司都非常重视结构数据库的开发和研究,在国内,开发自己的结构数据库尤其是中药有效成分三维结构数据库方面的工作进行得还很少,开发中药有效成分三维结构数据库对于中医药现代化的研究,从分子水平出发去研究中药以及中药新药的开发都具有非常重要的意义,我们课题组经过多年的努力,独立开发了包含6500个中药有效成分的三维结构数据库(3D-MSDT, 3D Molecular Structure Database of Traditional Chinese Drug),并配套开发了数据库检索和维护系统以及二维和三维分子显示系统,整个数据库系统既可单独运行,也可和PKUDDS^[1-3]实现无缝连结。

1 数据库及其软件系统的建立

中药有效成分的相关数据主要从报道的文献中采集,文献的范围包括重要的有关中草药有效成分的中医药典籍以及十余种与中药有关的中文杂志,从这些文献中,可以得到这些中药植物的不同用药部位所包含的有效成分分子结构、异名、化学名、物理性状、植物来源、药理药效、文献出处等等,把这些信息摘录并作成卡片,以便录入和以后的查验工作,对所有的结构,我们同时提供了二维结构和三维结构,所有的三维结构均采用MMFF力场进行分子力学优化^[4],优化工作在Cerius²分子模拟软件包中进行^[5]。

对于数据库中的数据,需要一套数据库管理系统来进行数据库的管理维护和数据的查询,国内已有的三维结构数据库一般采用商用ISIS软件系统^[6]。为了保持工作的相对独立性和创新性,我们独立开发了数据库管理和查询软件系统,这部分内容涉及大量的编程工作,包括下面的步骤。

1.1 3D-MSDT的后端支持和前端开发

3D-MSDT的后端支持选用了关系数据库模型进行数据的描述和管理,具体的数据库管理支持系统则选用Microsoft的SQL Server,SQL Server能够解决在Windows NT平台上的复杂问题,它可以支持ODBC,ADO,ANSI-SQL,HTML,Java和大部分主要的数据库访问和通讯协议,它还有助于我们有效和方便地开发丰富的应用程序,一系列多种多样的开发工具(如Visual C++, Visual Basic, ASP)可以在多种接口上进行定制。

3D-MSDT以SQL Server为数据库管理系统建立了3个库用于存放中草药的结构和性质数据,2个库用于辅助数据库前端程序的检索,存放临时信息,DrugStruDB库用于存放中草药有效成分的所有结构有关的信息,DrugPropDB库用于存放中草药有效成分的所有与结构无关的性质信息,而DrugNameDB库用于存放中草药有效成分的化合物名称信息,表1中显示了DrugStruDB的库结构,表中Stru3D字段存放着分子的三维结构,以mol文件格式储存;Image2D字段存放该化合物的标准二维结构图,以WMF文件的格式储存;AdjTable字段存放分子的连接表信息,这里的连接表是经过处理的,为C++一个结构数组的内存映像,这样可以加速子结构匹配时的搜索速度。

表 1 DrugStruDB 的库结构

Table 1 The database structure of DrugStruDB

字段名	数据类型	长度	精度	程度	是否可为空	缺省值	是否主键
ID	int	4	10	0	N	99999	Y
Formula	varchar	50	0	0	N	0	N
Weight	decimal	5	9	4	N	0.0	N
WeightText	char	10	0	0	N	0	N
AtomNo	int	4	10	0	N	0	N
AtomC	int	4	10	0	N	0	N
AtomH	int	4	10	0	N	0	N
AtomO	int	4	10	0	N	0	N
AtomN	int	4	10	0	N	0	N
AtomP	int	4	10	0	N	0	N
AtomS	int	4	10	0	N	0	N
Stru3D	text	16	0	0	Y	-	N
AdjTable	image	16	0	0	Y	-	N
Image2D	image	16	0	0	Y	-	N

我们根据数据库组织的三层模型,使用一个视图将 DrugStruDB 和 DrugPropDB 结合成为一个虚拟的“数据库”。这样的好处在于隐藏了数据的物理结构,表现在使用者面前的是一个数据的逻辑结构,这样使得数据库的扩充和修改不容易影响已开发的数据库应用,使得数据库对于应用程序来说具有良好的向下兼容性。

为了更好的与我们的 PKUDDS 系统无缝集成,3D-MSDT 采用 Visual C++ 6.0 为前端开发工具。在 Visual C++ 中,我们选用了 Visual C++ 中最先提供的面向对象的 OLE DB 技术 ADO,这样便于以后的扩展以及向网页查询迁移。

1.2 数据库的检索系统

3D-MSDT 的检索系统可以分成两个部分,一是在中草药有效成分库中进行的名称、分子式、活性等简单字符串匹配检索;二是带有初步药效团搜索色彩的子结构匹配检索(包含某些原子为任意原子的模糊检索)。这两个部分在 3D-MSDT 中的实现是不同的,第一部分主要涉及数据库 SQL 查询语言的应用,第二部分则侧重于算法的选择和实现。

1.2.1 结构性质的简单检索 对于 3D-MSDT 数据库中的凡是关于这些字段的信息都可以通过普通数据库查询语言 SQL 查询,但是根据这些字段在检索中可能遇到的几率,我们挑选了一些字段设计了友好的用户界面,使用户不用了解 SQL 语言就可以按照自己的需要进行检索,但是同时提供了一个专家模式可以绕过用户界面直接使用 SQL 语言更自

由的对所有字段进行组合检索。



图 1 系统的检索界面

Fig.1 The query interface of the system

在简单检索中,我们提供了分子量、分子式等结构检索和名称、原植物、生理药理相关、CAS 登记号等性质检索。在检索范围上,我们提供了选项,用户可以确定本次检索是在上一次检索的基础上(也即是上一次的检索结果集中)检索还是在整个数据库中检索。这样就可以使用多个检索条件快速逼近需要的结果集。同时系统提供了一次 Undo 的机会,这样可以避免误操作造成的影响。图 1 显示了系统的主检索界面。

1.2.2 子结构匹配检索 二维子结构匹配检索是我们数据库系统中最重要的一种查询方式。考虑到

中草药三维结构数据库的规模以及算法的可靠性,我们采用了 Xu 等人提出的基于偏序理论的回溯算法——GMA 算法^[7]。在子结构匹配中,对于 mol 文件格式,采用了自己的邻接表描述方式——一个由用于描述节点原子的结构所组成的数组。将得到的每一个邻接表直接以数组的二进制形式的内存映像存入数据库的 AdjTable 字段中,在执行子结构匹配程序代码之前,从数据库中读出邻接表后,于内存中分配该邻接表所需大小的内存,不需任何转换就可直接把读出的邻接表存入该内存空间,将内存指针转换后就得到了邻接表数组,这样匹配的速度可以加快。

子结构检索的输入是采用 mol 文件描述子结构作为检索算法的输入,在任何能够产生 mol 文件的分子构造程序中画出要检索的子结构即可。3D-MSDT 中内置了一个分子二维结构构造模块,可以非常方便及时地画出要检索的子结构。图 2 显示了在内置的分子构造模块中构造子结构。

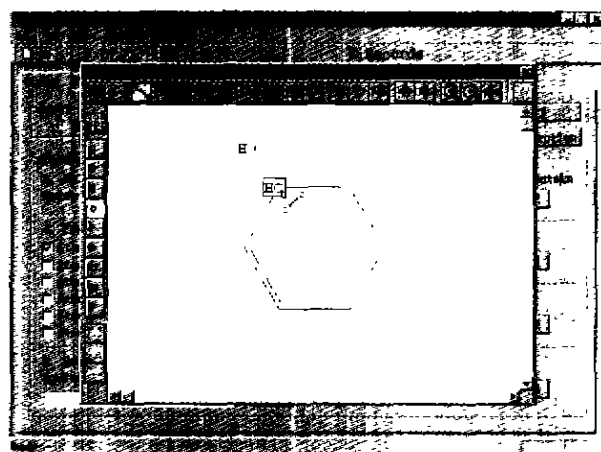


图 2 系统内置的子结构构造模块

Fig.2 The embedded substructure modeling module in the system

1.3 检索结果的浏览和输出

当用户提交检索条件后,3D-MSDT 会快速找到符合条件的化合物。如果结果集非空,系统会自动由检索页面跳转到浏览页面对结果集进行观察,图 3 即是以“水飞蓟”为检索词在原植物检索中得到的浏览页面。

浏览页面的右下角为分子的二维结构图像,还有一个“Show 3D”的按钮,按下可以在 PKUDDS 中观察这个分子的三维空间结构,如图 3 所示。上端工具条提供了浏览所需的按钮,可以向后、向前浏览结果集中的化合物,也可以直接跳到某一条记录。

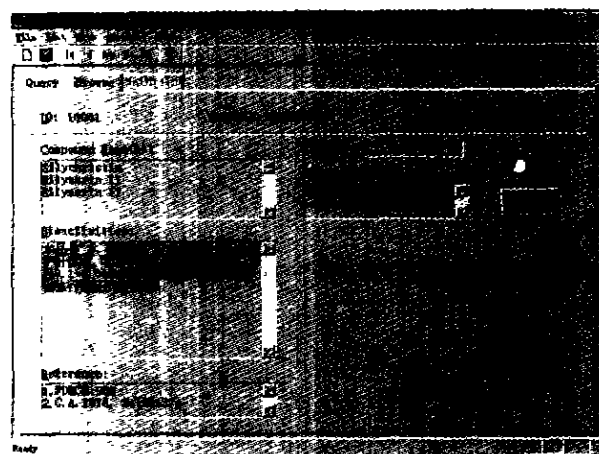


图 3 检索结果的显示

Fig.3 The presentation of the searching results



图 4 在 PKUDDS 中分子三维结构的显示

Fig.4 The 3D molecular display in PKUDDS

为了与其他数据库互换数据,3D-MSDT 支持把检索集中的所有化合物结构信息以 SD 文件的形式存盘。绝大多数商业化学数据库搜索软件可以使用 SD 文件输入数据库,数据的可交换性增加了 3D-MSDT 的扩展性。

1.4 数据库的数据的维护

作为一个完善的数据库系统,3D-MSDT 不仅有数据的检索和浏览的能力,还有完善的数据维护功能。通过友好的用户界面,用户不用了解底层的 SQL 语言细节即可管理数据,对 3D-MSDT 进行数据的增添、删除和更改。

3D-MSDT 的数据维护界面和浏览界面基本近似,只是文本框的属性由只读变为可写,方便更改数据。图 5 为新增一个化合物并指定三维结构文件后的系统图示,可以看到 ID、分子式和分子量都已经自动产生。为了适应大数据量的输入输出维护,3D-MSDT 提供了批处理模式。点击“Batch”按钮,即

可进入批处理对话框,这时只需指定起始和终止的ID号,其它信息事先输入并存放于磁盘文件中,即可批量更新,也可批量导出数据以供它用,提供了“追加(Append)”和“更新(Refill)”两种批处理模式,分别适应于新增记录和将原有记录导出并修改后的更新。

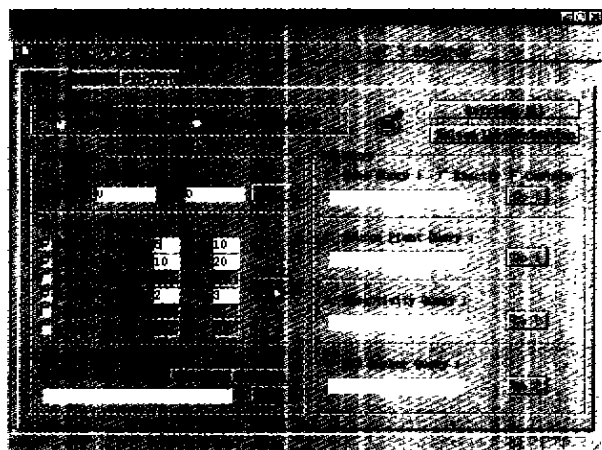


图5 数据库系统的维护界面

Fig.5 The maintenance interface of database system

2 数据库系统的特色

同目前已有的国外结构库及国内已有的为数不多的天然产物数据库相比^[8],我们的中药有效成分三维结构数据库具有显著的特色.首先,数据可靠性强,目前所知的国内外已有的大型三维结构数据库中的结构都采用三维结构产生程序从二维结构直接生成,这种方法虽然速度快,但转化而来的结构在结果准确性和结构手性上存在明显的缺陷,从而使数据库的可靠性大打折扣,我们在建库过程中,所有的三维结构都不通过二维结构的转化方法,而是用三维结构直接优化的方法,所有结构都通过分子力学优化,优化采用了 MMFF 力场,并且在优化过程中考虑了静电相互作用,优化在 Cerius² 分子模拟软件包上进行^[5].考虑到分子的手性问题直接关系到其活性,我们特别注意了分子的构型问题,这样就保证了后续的药物设计过程中小分子三维结构的准确性。

我们采用的是独立开发的数据库检索和管理维护系统,而不是可购置的一般商业软件,这是国内大部分数据库所不具备的。

系统另外一个重要特点就是实现了该数据库和 PKUDDS 软件的无缝结合,在数据库系统中,直接采用了 PKUDDS 作为三维结构显示模块,检索得到的分子可以直接从数据库系统中调入 PKUDDS.在 PKUDDS 中,分子可以以多种形式显示,而且可以利用 PKUDDS 中的软对接和柔性对接方法进行基于受体结构的药物分子的虚拟筛选^[9-11]。

该数据库系统和我们科研组开发的北京大学药物设计系统以及中草药信息系统构成了完整的基于中药的药物设计系统^[12],该系统已经用于 NS3-NS4A 蛋白酶抑制剂以及其它体系的研究并取得了很好的结果。

本文为“庆祝邢其毅教授九十华诞暨执教六十年”征文

References

- Hou, T. J.; Xu, X. J. *J. Mol. Graphics Modell.*, **2001**, *19*, 455.
- Xu, X. J.; Hou, T. J.; Wang, J. M.; Liao, N.; Luo, H. P.; Gao, S. L. *Computer and Applied Chemistry*, **1999**, *16*, 335 (in Chinese).
- Xu, X. J. *Progress in Chemistry*, **1999**, *11*, 202 (in Chinese).
- Hagren, T. A. *J. Comput. Chem.*, **1996**, *17*, 490.
- Cerius² User Guide, Molecular Simulations, INC., San Diego, USA, **1998**.
- INS Base Database Maintenance, MDL Information Systems, INC., San Leandro, USA, **1999**.
- Xu, J. J. *Chem. Inf. Comput. Sci.*, **1996**, *36*, 25.
- Chapman and Hall Natural Product, Tripos, Inc., Louis, USA, **2001**.
- Wang, J. M.; Hou, T. J.; Xu, X. J. *Chemom. Intell. Lab. Syst.*, **1999**, *45*(1-2), 281.
- Hou, T. J.; Wang, J. M.; Xu, X. J. *Protein Eng.*, **1999**, *12*, 630.
- Hou, T. J.; Wang, J. M.; Xu, X. J. *Chin. Chem. Lett.*, **1999**, *30*, 615.
- Yu, H. D.; Hou, T. J.; Qiao, X. B.; Xu, X. J. *Chinese Journal of Integrated Traditional and Western Medicine*, submitted (in Chinese).

(Ed. CHENG Biao)

(DONG Hua-Zhen)