

海洋生物及其代谢产物管理系统的设计与实现*

胡谷平^{1,2}, 孙丽^{1,3}, 陈省平^{1,3}, 袁洁^{1,3}, 蓝秀健^{1,3},
吴珏珩^{1,3}, 余志刚^{1,2}, 林永成^{1,2}

1. 中山大学海洋微生物功能分子广东高校重点实验室, 广东 广州 510275;
2. 中山大学化学与化学工程学院, 广东 广州 510275;
3. 中山大学中山医学院, 广东 广州 510089)

摘要: 对海洋产物及其代谢产物的信息进行了分析, 以化学类、生物类和药学类权威杂志为数据库信息来源, 设计并建立了海洋生物及其代谢产物数据库 (Marine Organism Metabolite Data Center, MOMDC)。MOMDC 是一个包括海洋生物、代谢产物、生物学活性三方面信息的公益性网络数据库, 目前已收录的信息为 19 000 条。

关键词: 海洋生物; 代谢产物; 数据库

中图分类号: Q8H.4 文献标志码: A 文章编号: 0529-6579(2010)06-0089-04

Design and Implementation of the Database Management Marine Organism Metabolite

HU Guping^{1,2}, SUN Li^{1,3}, CHEN Shengping^{1,3}, YUAN Jie^{1,3}, LAN Jianxiu^{1,3},
WU Jueheng^{1,3}, SHE Zhigang^{1,2}, LIN Yongcheng^{1,2}

1. Functional molecules of Marine Micro-Organisms, Sun Yat-sen University, Guangdong Key Laboratory of Universities, Guangzhou 510275, China;
2. School of Chemistry and Chemical Engineering, Sun Yat-sen University, Guangzhou 510089, China;
3. Zhongshan School of Medicine, Sun Yat-sen University, Guangzhou 510275, China)

Abstract: By analyzing the information from marine organisms and their metabolic products, Marine Organism Metabolite Data Center is designed which the information source is from the chemical, biological and pharmacy authority magazine. MOMDC is a nonprofit network database that includes marine biological, metabolic products and biological activity, that consist of about 19 000 marine information.

Key words: marine biological; metabolic product; database

海洋占地表面积的 71%, 是一个巨大的资源宝库。海洋环境中包括了高盐、高压、低温、低营养、无光照等极端生存环境。因此海洋生物大多具有一些特殊的结构和功能, 以维持其生命活动。同时, 海洋生物的多样性、复杂性和特殊性使源于其中的代谢产物也具有多样性、复杂性和特殊性。近年来, 随着科学技术的发展, 人们不仅发现了一系列新的海洋生物物种, 更重要的是从海洋生物代谢

产物中分离得到了大量的新结构化合物, 为寻找新的药用活性物质提供了丰富的源泉^[1]。据我们统计, 近 20 年来从海洋中发现的新结构的化合物超过 10 000 个, 并且呈几何级数的趋势增长。这些产物大多有重要的生物活性, 尤其表现在抗癌活性、抗真菌活性、对心血管作用的活性上^[2-3]。而这些新结构的化合物的发现, 也成为了世界药物开发的热点^[4]。

面对如此海量的海洋生物和代谢产物, 如何更

* 收稿日期: 2010-01-06

基金项目: 国家海洋 863 计划资助项目 (2007AA09Z431); 国家科技基础性工作专项资助项目 (2007FY210600); 广东省重大科技专项资助项目 (2007A032600001)

作者简介: 胡谷平 (1974 年生), 男, 实验师, 博士; 通讯作者: 陈省平;; E-mail: chenshp@mail.sysu.edu.cn

好的交流和共享是一个迫切需要解决的问题^[5]。海洋生物及其代谢产物数据库 (Marine Organism Metabolite Data Center, 简称 MOMDC) 就是基于这个目的建立的网络信息平台。该平台通过收集海洋生物及其代谢产物的信息, 建立一个包括海洋生物、代谢产物、生物学活性三方面的网络数据库, 实现信息共享, 为海洋生物及其代谢产物的研究和新药的开发提供信息支持^[6]。

2 MOMDC 数据库的设计

2.1 数据来源

MOMDC 收集了近 30 年来世界范围发现的新的海洋生物及其代谢产物的信息, 到目前为止, 已收录的海洋生物信息为 5 283 条, 代谢产物信息为 9 608 条, 同时从海洋生物信息和代谢产物信息中分析整理出生物活性信息共 4 060 条。这些信息来自 110 种学术期刊, 笔者对这些期刊出现的频率进行了统计, 表 1 列出了各信息中排名前 10 的杂志。

表 1 MOMDC 中引用期刊情况 (引用超过 10 条)

Table 1 the journal in the MOMDC (which referenced more than 10 times)

化学类	生物类	药学类
J Nat Prod.	Chem Biodirvers	J Antibiot
Tetrahedron	Chemosphere	Med Res Rev
J Org Chem	J boil chem	Mar Drugs
Tetrahedron Lett	Biosci Biotechnol Biochem	J Anti Biot
J Am Chem Soc	Int J Sys Evo Mic	Biol Pharm Bull
Phytochemistry	Appl Microbiol	Pharm Biol
Chem Lett	J Environ Monit	Curr Drug Metab
Can J Chem	J Agric Food Chem	Toxicon
Org Lett	Bioorg Med Chem	Toxicology
Chem Commun	Cryobiology	Aquat Toxicol

从上表中可以看出, MOMDC 囊括了化学、生物和药学三方面的信息, 保证了数据的全面性; 同时以各学科的权威期刊作为信息的数据来源, 保证了数据的可靠性。

2.2 数据库的结构

MOMDC 的数据分别来自海洋生物和海洋生物代谢产物两个方面, 但考虑到这两个方面的研究工作主要是为新药研发做准备, 因此我们将生物活性信息, 单独提取出来成为一个独立组成部分^[7]。表 2-4, 是 MOMDC 设置的 3 个主表, 分别记录海洋生物信息、海洋生物代谢产物信息和相关的生物学活性信息。

表 2 海洋生物信息主表 (MO-info)

Table 2 information sheet of marine organisms

编号	关键词	内容	数据类型
1	MOid	海洋生物信息数据库的唯一标识符	数字型
2	MOcs	海洋生物保藏地点	字符型
3	MOgenus	海洋生物属名	字符型
4	MOstrian	海洋生物种名	字符型
5	MOcollp	海洋生物采集地点描述	字符型
6	MOcollt	海洋生物采集时间描述	字符型
7	MOgs	海洋生物生长的基物	字符型
8	MOs	海洋微生物的培养基	字符型
9	MOstemp	海洋微生物的培养温度	字符型
10	MOmeta	海洋生物各种代谢产物	字符型
11	MOcolle	海洋生物采集地点的生长环境	字符型
12	MOspec	海洋生物本身的各种特殊的生长特点、形态特点	字符型
13	MOpic	海洋生物的照片包括显微照片或普通照片	字符型
14	MOref	海洋生物的参考文献	字符型

表 3 海洋生物代谢产物信息主表 (MOM-info)

Table 3 information sheet of marine organisms Metabolite

编号	关键词	内容	数据类型
1	MOMid	代谢产物在数据库中的唯一标识	数字型
2	MOMname	代谢产物的系统命名或俗名	字符型
3	MOMf	代谢产物的分子式	字符型
4	MOMfw	代谢产物的相对分子质量	字符型
5	MOMaa	代谢产物的元素组成	字符型
6	MOMfrom	代谢产物是从哪一个海洋生物中分离出来的	字符型
7	MOMl	代谢产物结构相似的化合物	字符型
8	MOMs	代谢产物化学结构中主要官能团和结构特点	字符型
9	MOMst	代谢产物的外观性状, 包括形态、颜色等	字符型
10	MOMmp	代谢产物的熔点	字符型
11	MOMp	代谢产物的荧光光度, 包括测试温度和光源等信息	字符型
12	MOMcd	代谢产物的圆二色谱	字符型
13	MOMuv	代谢产物的紫外光谱, 包括吸收峰位置和强度	字符型
14	MOMir	代谢产物的红外光谱, 包括波数和强度	字符型
15	MOMhnmr	代谢产物的氢谱, 包括溶剂、仪器兆数, 吸收峰位置、峰型、耦合常数等	字符型
16	MOMcnmr	代谢产物的碳谱, 包括选用溶剂、仪器兆数, 吸收峰位置	字符型
17	MOMms	代谢产物的质谱, 包括电离方式, 分子离子峰位置等, 尽量提供高分辨质谱	字符型
18	MOMpha	代谢产物的各种生物活性	字符型
19	MOMs	代谢产物的是否已经被合成	字符型
20	MOMstruct	代谢产物的化学结构图	二进制型
21	MOMref	代谢产物的相关参考文献	字符型

表4 生物学活性信息表 (MOMphainfo)

Table 4 information sheet of marine organisms Bioactive

编号	关键词	内容	数据类型
1	MOid/MOMid	对应的海洋生物或代谢产物编号	数字型
2	PHAt	活性研究时的实验系统	字符型
3	PHAj	预期活性归类	字符型
4	PHAn	生物学活性效应值	字符型

MOMDC 是一个典型的多学科交叉数据库，海洋生物信息表中的代谢产物数据域，可以关联到海洋生物代谢产物信息表，同时海洋生物代谢产物信息表中的代谢产物来源数据域也可以关联到海洋生物信息表。生物学活性信息表中也包含了具有这些生物学活性的海洋生物和代谢产物，三信息主表的关系如图1所示。通过这3个主表的交叉检索，构成一个完整的数据关联体系，有效地整合三方面的信息资源。

2.3 数据库的管理系统和网络搜索平台

MOMDC 采用三层架构设计 (3-tier application)。即工作站表示层接口，事务逻辑层和数据库及其相关程序设计层^[8]。三层架构可以更好的支持分布式计算环境。逻辑层的应用程序可以在多个机器上运行，充分利用网络的计算功能。此外，

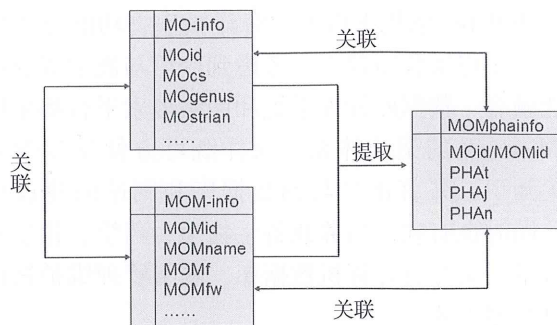


图1 MOMDC 中三主表关系图

Fig. 1 The relationship of MOMDC's sheets

三层架构的最大优点是其安全性。用户端只能通过逻辑层来访问数据层，减少了入口点，屏蔽了多数危险的系统功能^[9]。

MOMDC 的数据库管理系统使用的是 Microsoft SQL Server 2003，网络服务器采用 Win2003 Server + IIS6.0 架构，网络界面和后台搜索程序采用 ASP.NET 2.0 技术构建，脚本语言使用 Java Script，保证了程序的兼容性。同时，MOMDC 还会设立一个网上学术交流论坛，为相关研究工作者提供一个学术交流的平台，并且可以通过网上的讨论为 MOMDC 的发展提供好的意见和建议^[10]。由此便搭建出图2所示的数据库总体运行图。

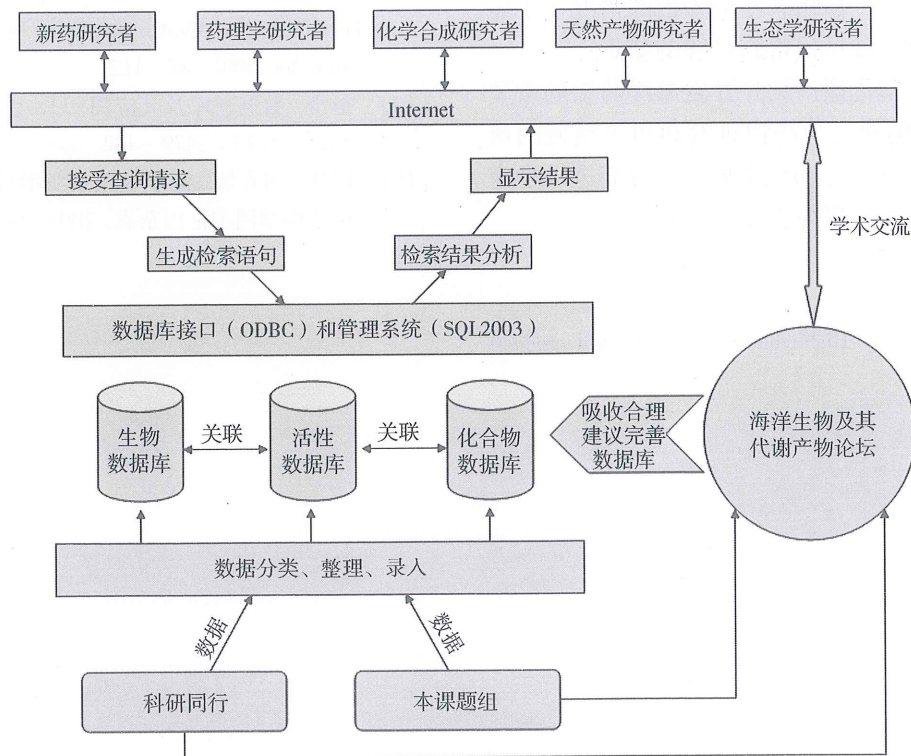


图2 网站的运行模式图

Fig. 2 The flowchart of web site

MOMDC 依托中山大学海洋微生物功能分子广东高校重点实验室建立, 考虑到今后与教学资源的可能融合, 数据库管理系统和网络搜索平台都采用了 Microsoft 的架构体系。这样既可方便学生学习相关知识, 并直接参与到数据库和网站的建设中来, 同时也有利于培养兼备生物学、药学、化学专业背景, 又具有计算机数据库、网络管理维护知识的复合型人才。

3 总 结

进入 21 世纪以来, 互联网的大规模普及和发展, 改变了依靠纸质期刊的传统科学交流模式。MOMDC 正是一个建立在互联网上的信息交换平台, 它具有以下 3 个特点: ①综合性数据库。MOMDC 中包含了海洋生物的生物信息、代谢产物的化学信息和生物学活性信息, 实现关联交互检索, 为来自生物学、化学、药学的学者提供了一个信息交换平台。②信息新颖全面, 近 20 年来, 海洋中发现的新化合物一直是新药开发的重点。MOMDC 收录的信息就是关于这些新化合物和产生它们的海洋生物。这些信息的汇集, 形成了一个海洋生物及其代谢产物的信息交流平台, 为进一步开发利用这些资源提供了方便。③事实型数据库^[11]。对于每一个代谢产物和海洋生物, MOMDC 都收录了较为全面的信息, 使用者在大多数的情况下不需要查找原始文献, 就可以得到足够的信息。

海洋生物及其代谢产物是开发海洋药物的重要来源。MOMDC 的建立为药物研发和相关领域的研究人员提供了信息支持和交流平台。我们进一步完善和扩张数据库, 为研究者提供便利。

参考文献:

[1] FAULKNER D J. Highlights of marine natural products

chemistry (1972 - 1999) [J]. *Natural Product Reports*, 2000, 17: 1 - 6.

- [2] PIGGOTT A M, KARUSO P. Quality not Quantity: The role of marine natural products in drug discovery and reverse chemical proteomics [J]. *Mar Drugs*, 2005 (3): 36 - 63.
- [3] JHA R K, XU Z R. Biomedical Compounds from Marine organisms [J] *Mar Drugs*, 2004 (2): 123 - 146.
- [4] FUSEYANI, NOBUHIRO. Biotechnological potential of marine natural products [J]. *Pure and Applied Chemistry*, 2010, 82 (1): 17 - 26.
- [5] 刘涛, 陈省平, 汪小龙, 等. 海藻种质管理信息系统的开发与应用 [J]. *高技术通讯*, 2004 (6): 83 - 84.
- [6] BRUNDEGE J M, CHRISTOPHER D. Baroquely: an object framework for building queries to biomedical databases [J]. *Bioinformatics*, 2003, 19 (7): 901 - 902.
- [7] PROKSCH P, EDRADA R A, EBEL R. Drugs from the seas - current status and microbiological implications [J]. *Appl Microbiol Biotechnol*, 2002, 59: 125 - 134.
- [8] CORBIN R D, DUNBAR C B, ZHU Q M. A three-tier knowledge management scheme for software engineering support and innovation [J]. *The Journal of Systems and Software*, 2007, 80 (9): 1494 - 1505.
- [9] MALHOTRA Y. Why knowledge management systems fail? Enablers and constraints of knowledge management in human enterprises [C] // MICHAEL E D K, SRIKANTIAH T K, eds. *Knowledge Management Lessons Learned: What Works and What Doesn't*. Information Today Inc, 2004: 87 - 112.
- [10] 彭涛. 数据仓库及其应用 [J]. *计算机与应用化学*, 2003, 20 (4): 479 - 482.
- [11] 何玮, 雷春炳. 医学专题事实型数据库的探讨 [J]. *中华医学图书情报杂志*, 2002, 11 (2): 3 - 6.