mammalian embryos. TIG, 1996, **12** (4): 134~ 138
15 Szabo P E, Mann J R. Allele-specific expression and total expression levels of imprinted genes during early mouse development-implications for the imprinting mechanism. Gene Develop, 1995, **9** (24): 3097~ 3108
16 Lerchner W, Barlow D P. Paternal repression of the imprinted mouse Igf2r/Mpr 300 locus occurs during implantation and is stable in all tissues of the post implantation mouse embryo. Mech Develop, 1997, **61** (1): 141~ 149
17 Lee J T, Jaenisch R. Long-range *cis* effects of ectopic X-inactivation centres on a mouse autosome. Nature, 1997, **386** (6622): 275~ 279
18 Herzing L B K, Romer J T, Horn J M, *et al*. Xist has properties of the X-chromosome inactivation center. Nature, 1997, **386** (6622): 272~ 275
19 Kim J, Ashworth L, Branscomb E, *et al*. The human homolog of a mouse-imprinted gene, Peg3, maps to a zinc finger gene-rich region of human chromosome 19q13. 4. Genome Res, 1997, **7** (5): 532~ 540
20 Beechey C V, Cattanach B M. Genetic and physical imprinting map of the mouse. Mouse Genome, 1997, **95** (1): 100~ 105
21 McGrath J, Solter D. Completion of mouse embryogenesis requires both the maternal and paternal genomes. Cell, 1984, **37** (1): 179~ 183
22 Surani M A H, Barton S C, Norris M L. Development of reconstituted mouse eggs suggests imprinting of the genome during gametogenesis. Nature, 1984, **308** (5959): 548~ 550
23 Ledbetter D H, Engel E. Uniparental disomy in humans-development of an imprinting map and its implications for prenatal diagnosis. Hum Molecular Genet, 1995, **4**: 1757~ 1764
24 Guillemot F, Caspary T, Tilghman S M, *et al*. Genomic imprinting of Mash2, a mouse gene required for trophoblast development. Nat genet, 1995, **9** (3): 235~ 242

Imprinted Genes and Embryos. LI Hong-Jun, GUO Ying-Lu (*Institute of Urology, Beijing Medical University, Beijing* 100034, China); ZHANG Zhi-Wen (*Department of Physiology, Beijing Medical University, Beijing* 100083, China).

Abstract Some genes express only one allelic gene and the another allelic gene is specifically inhibited through some kinds of gene modification. Those genes are called imprinted genes, and they are a unique model of allelic exclusion. Most of imprinted genes participate in regulating development and differentiation of embryo and newborn infant, and disorder of imprinting function may results in many kinds of abnormal development and stilbirth. The mechanisms for the formation of imprinted genes, specific recognization and the defect of imprinting function are still not very clear.

Key words imprinted gene, embryo, development

国内外生物信息学数据库服务新进展*

李维忠 王任小 林大威 毛凤楼 韩玉真 来鲁华[1)]
(北京大学物理化学研究所, 北京 100871)

摘要 生物信息学是生命科学中最活跃的领域之一. 各类生物信息学数据库在近年不断出现, 其规模呈爆炸趋势增长, 同时数据结构日趋复杂. 目前生物信息学数据库服务已实现了高度的计算机和网络化. 算法和软件的进步、数据库的一体化、服务器-客户模式的建立使之成为生物、医药、农业等学科的强有力工具. 在国内北京大学物理化学研究所于1996年建立了第一家生物信息学网络服务器. 现已为国内外科学家提供了7万余次服务, 在国际上具有一定影响.

关键词 生物信息学, 数据库, 软件
学科分类号 Q71

生物信息学 (bioinformatics) 是近年来发展并完善起来的热门交叉学科. 近年来随着快速序列测定、基因重组、多维核磁共振、同步辐射、机器人等技术的应用, 生物学实验数据呈爆炸趋势增长, 同时计算机和国际互联网络的发展使对大规模数据的存贮、处理和传输成为可能. 现在某一实验室的研究成果一经进入生物信息网络便为全球科学家共享. 从新基因的发现, 蛋白质的结构功能预测、疫苗的筛选到新药研制无不依赖于生物信息学, 它在生物、医药、农业、环境等学科的应用已无所不在.

* 863高技术计划 (960311280)、攀登计划——生命过程中重要化学问题研究 (970211006)、国家杰出青年科学基金 (29525306) 及国家教委资助. 1) 通讯联系人.
收稿日期: 1997-09-26, 修回日期: 1998-02-19

1 生物信息学数据库的发展现状

数据库是生物信息学的主要内容，各种数据库几乎覆盖了生命科学的各个领域，核酸序列数据库有 GenBank[1]，EMBL[2]，DDBJ[3] 等，蛋白质序列数据库有 SWISS-PROT[4]，PIR[5]，OWL，NRL3D，TrEMBL 等，蛋白质片段数据库有 PROSITE[6]，BLOCKS[7]，PRINTS 等，三维结构数据库有 PDB[8]，NDB，BioMagResBank，CCSD 等，与蛋白质结构有关的数据库还有 SCOP[9]，CATH，FSSP[10]，3D-ALI，DSSP 等，与基因组有关的数据库有 ESTdb，OMIM，GDB，GSDB 等，文献数据库有 Medline，Uncover 等. 此外还有其他数据库数百种. 另外一些公司还开发了商业数据库如 MDL 等.

1.1 数据库的增长和更新

爆炸性增长是生物信息学数据库的重要特征，至 1997 年底 GenBank 已有 189.2 万条核酸序列，SWISS-PROT 有 69 000 条蛋白质序列，PDB 有 7 000套结构. 目前这种趋势主要是因为基因组等计划的实施，一些物种（如酵母 *Saccharomyces cerevisiae*）基因组的全部序列已经收入 GenBank，在 GenBank 中有 65% 的序列来自 ESTs（Expressed Sequence Tags），预计人类基因组计划在 2 年内还将测出 10^8 核苷酸的序列.

1.2 数据库的复杂程度增加

除了在数量上的增长，数据库的复杂程度在不断增加. 它包括了大量注释、参考文献及软件，并通过指针将相关内容链接到其他数据库，在第 35 版 SWISS-PROT 中注释项涉及蛋白质的功能、结构域和活性位点、二级结构、四级结构、翻译后修饰、与其他蛋白质的相似性、相关疾病、序列冲突等. 与之交叉引用的数据库增加到 26 个. 数据库结构层次的加深客观上要求管理的进步，当今面对对象数据库管理方法正在逐步取代旧的模式.

1.3 数据库使用的高度计算机和网络化

计算机和网络化是生物信息学的又一重要特点. EMBNet（European Molecular Biology Network）已经连接了 22 个国家节点和 8 个大型生物计算中心. 成为最大的生物信息学区域网络. 许多数据库服务器已从工作站升级到大型服务器. 软件的进步使数据库实现了更为高效灵活的管理和使用. 1997 年 7 月 PDB 推出了自动数据投送系统 AutoDep，使向 PDB 投送数据操作大大减化. 而 EBI（European Bioinformatics Institute）则通过为一些机构建立帐户的方式提高数据收集的效率. JAVA 是一种不依赖于平台的高效网络语言，PRINTS 数据库在 1996 年就将 JAVA 引入，实现了交互式序列对比的可视化功能，而 PDB 等已经实现了 VRML（virtual reality modeling language）模型的传送，使用户可以在空间任一视点观察生物大分子的结构.

2 生物信息学网络上的数据库服务进展

生物信息学网络资源是以数据库为主体，包括软件、信息查询、专题组和公告牌、自动计算等多种工具的综合资源（表 1，详细资料见北京大学生物信息学服务器）.

2.1 生物信息学软件的进步

生物信息学各个领域中的软件数目庞大，在 EBI 1997 年的分子生物学程序目录中就收录了 530 多种常用软件. 序列对比和数据库搜索软件有 BLAST，FASTA，BLITZ 等，生物大分子可视化软件有 Rasmol，Mage，Raster3d，Grasp 等，与蛋白质结构有关的程序有 Procheck，WHATIF，DSSP 等，大型分子生物学软件包如 GCG. 各个数据库还有自身的查询系统. 并行算法、遗传算法、面向对象算法等已被应用到最新的程序中. 于 1996 年 9 月推出的 FASTA3 实现了高度的并行化，能够在多 CPU 的计算机上将搜索速度提高数倍，还支持并行虚拟机器技术（Parallel Virtual Machine），增加了并行网络计算机制.

2.2 数据库的一体化和集成环境

生物信息学数据库覆盖面广，分布分散且格式不统一，因此一些生物计算中心将多个数据库整合在一起提供综合服务. EBI 的 SRS（sequence retrieval system）包含了核酸序列库、蛋白质序列库，三维结构库、基因组等 30 多个数据库及 FASTA、CLUSTALW、PROSITESEARCH 等强有力的搜索工具，用户可以进行多个数据库的多种查询. 1997 年升级的 PDB 3DBBrowser 搜索软件，可以接受各种关键词的查询，还具备字典功能和 FASTA 序列搜索功能. 用户不仅可以得到生物大分子的各种注释、坐标、三维图形，VRML 等，并能从一系列指针连接到 SCOP、CATH、Medline、ENZYME、Swiss-3Dimage 等.

表 1 部分重要的生物信息学网络资源

数据库及服务	网址
EMBL	http://www.embl-heidelberg.de/
GenBank	http://www.ncbi.nlm.nih.gov/Web/Genbank/
DDBJ	http://www.ddbj.nig.ac.jp/
SWISS-PROT	http://www.expasy.ch/sprot-top.html
PIR	http://www-nbrf.georgetown.edu/pir/
GDB	http://gdbwww.gdb.org/
PDB	http://www.ipc.pku.edu.cn/npdb/
SCOP	http://www.ipc.pku.edu.cn/scop/
EBI	http://www.ebi.ac.uk/
NCBI	http://www.ncbi.nlm.nih.gov/
ExPASy	http://www.expasy.ch/
SRS	http://srs.ebi.ac.uk:5000/
Entrez	http://www3.ncbi.nlm.nih.gov/Entrez/
Weizmann Institute	http://bioinformatics.weizmann.ac.il/
Pedro's BioMolecular Research tools	http://www.public.iastate.edu/~pedro/research_tools.html
Medline	http://www2.ncbi.nlm.nih.gov/medline/query_form.html
BioMedNet	http://www.BioMedNet.com/

2.3 服务器-客户式结构

生物信息网络中的数据库服务广泛采用服务器-客户式结构，这些服务器包括为数众多的数据库搜索和序列对比服务器以及各专业领域的服务器. 位于以色列 Weizmann 研究所的 Bioccelerator 是为序列搜索设计的专用并列计算机，它将 Swith-Waterman, BLAST, FASTA 等方法硬件化，实行并行计算和先进的内存管理，令搜索速度大幅度提高，是生物信息学服务器更新的一个重要实例. PredictProtein 是蛋白质结构预测服务器，它可根据要求的方法计算出所求蛋白质多重序列对比的结果、二级结构、残基可及性、跨膜螺旋位置、折叠拓扑类型等. DALI 是计算蛋白质折叠类型和三维结构对比服务器，用户输入蛋白质的结构，由服务器给出 PDB 中与之具有相似结构的蛋白质及叠合的参数.

3 国内进展及北京大学生物信息学服务器

近两年国际上生物信息学发展异常迅猛，为带动我国生物信息学的发展，北京大学物理化学研究所于 1996 年建立了国内第一家生物信息学网络服务器. 通过 WWW (http://www.ipc.pku.edu.cn), FTP (ftp://ftp.ipc.pku.edu.cn) 及 E-mail 方式为我国及世界各地科学家提供数据库、生物信息资源查询、软件和电子邮件等多种服务 (见表 2).

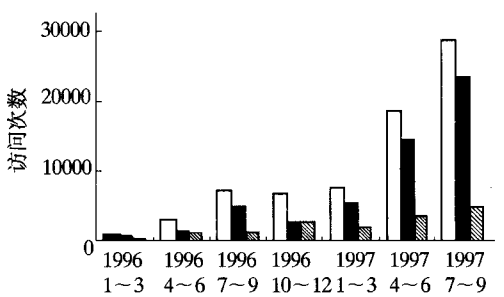
表 2 北京大学生物信息学服务器的服务内容

名称	内容	网络协议
PDB[1)]	Brookheaven Protein Data Bank	www, ftp
SCOP[1)]	Structural Classification of Proteins	www
SWISS-PROT	Protein Sequence Data Base	ftp
PIR	Protein Identification Resource	ftp
ENZYME	Enzyme Data Bank	ftp
PROSITE	Protein Sites and Patterns Database	ftp
BLOCKS	Database of Highly Conversed Regions in Proteins	ftp
DSSP	Database of Secondary Structure of Proteins	ftp
FSSP	Database of Families of Structurally Similiar Proteins	ftp
HSSP	Database of Homology-derived Structures of Proteins	ftp
Database Search	Database Search	www
Journal	On-line Bio/Chemical Journals	www
Course	Courses, Guides, Help and Tutorials	www, ftp
BioServices	Bioservices Search	www
Software Search	Software Search	www
Software Archie	Collection of Free Softwares	www, ftp

[1)] Mirror sites.

数据库服务包括 PDB 及 SCOP 的镜像服务器 (mirror) 及其他重要的数据库. 通过对镜像的访问，用户可以得到和原数据库完全相同并同步更新的服务. 信息资源查询包括数百个数据库及网址的查询、著名杂志及文献服务器的查询，此外服务器还提供遗传算法、蛋白质结构原理等训练课程. 服务器的软件服务包括各类生命科学软件的查询及可供用户下载常用免费软件的 FTP 服务，我们将本实验室开发的软件也置于网上以扩大国内工作的国际影响. 目前发布的程序有基于受体结构的药物分子设计程序 RASSE[12]、蛋白质局部区域计算程序 LPSA[13]、有机化合物脂水分配系数的计算程序 XLOGP[14]，这些程序在国际已经有数十家用户.

1996 年 5 月，本服务器正式发布并召开了国内首次生物信息学应用研讨班，国内有数十家单位参加，反映良好. 目前服务器接收到国内外70 000余次访问 (图 1). 在今后我们计划进一步加强服务功能，使之成为生物信息学学术交流的园地和对外宣传的窗口.



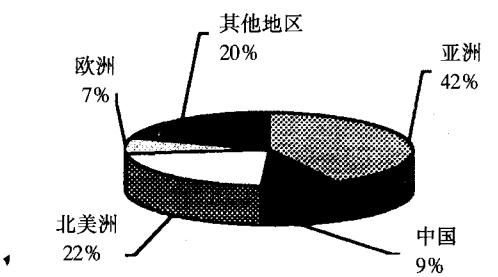


图 1　北京大学生物信息学服务器从 1996 年 1 月至 1997 年 9 月接受访问的次数及地区分布

□: ALL; ■: SCOP; ▨: PDB.

4　生物信息学数据库的研究应用前景

对于绝大多数科研工作者，生物信息学无疑是一种强有力的工具. 而对已有数据的研究和理解将一直是富有挑战性重要课题. 根据统计原理，某一程度上统计结果的显著性与数据量的对数成正比. 因此大量基于数据库的研究工作将随数据库容量的增长而有所突破. 如蛋白质的结构预测目前虽无法解决，但随着 PDB 中数据特别是新折叠类型的大量增加，此难题必会有重大进展. 人类基因组计划旨在从基因水平上提示疾病的本质，在读出全部基因组序列后，生物信息学最重要的研究内容就是如何读懂这些数据，从而揭示生命的奥秘. 生物信息学科也将随之得到巨大发展.

参　考　文　献

1　Benson D A, Boguski M S, Lipman D J, *et al*. GenBank. Nucleic Acid Res, 1997, **25** (1): 1~6

2　Stoesser G, Sterk P, Tuli M A, *et al*. The EMBL Nucleotide Sequence Database. Nucleic Acid Res, 1997, **25** (1): 7~13

3　Tateno Y, Gojobori T. DNA Data Bank of Japan in the age of information biology. Nucleic Acid Res, 1997, **25** (1): 14~17

4　Bairoch A, Apweiler R. The SWISS-PROT protein sequence data bank and its supplement TrEMBL. Nucleic Acid Res, 1997, **25** (1): 31~36

5　Sidman K E, George D G, Barker W C, *et al*. The protein identification resource (PIR). Nucleic Acid Res, 1988, **16** (5): 1869~1871

6　Bairoch A, Bucher P. PROSITE: recent development. Nucleic Acid Res, 1994, **22** (17): 3583~3589

7　Henidoff S, Henikoff J G. Automated assembly of protein blocks for database searching. Nucleic Acid Res, 1991, **19** (23): 6565~6572

8　Bernstein F C, Koetzle T F, Williams G J B, *et al*. The Protein Data Bank: a computer-based archival file for macromolecular structures. JMB, 1977, **112** (3): 535~542

9　Murzin A G, Brenner S E, Hubbard T, *et al*. Scop: a structural classification of proteins database for the investigation of sequences and structures. JMB, 1995, **247** (4): 536~540

10　Ho lm L, Sander C. Dali/FSSP classification of three-dimensional protein folds. Nucleic Acid Res, 1997, **25** (1): 231~234

11　Bairoch A. The ENZYME data bank. Nucleic Acid Res, 1994, **22** (17): 3626~3627

12　Luo Z, Wang R, Lai L. RASSE: a new method for structure-based drug design. J Chem Inf Comp Sci, 1996, **36** (6): 1187~1194

13　Zhang H, Lai L, Wang L, *et al*. A fast and efficient program for modeling protein loops. Biopolymers, 1997, **41** (1): 61~72

14　Wang R, Fu Y, Lai L. A new atom-additive method for calculating partition coefficients. J Chem Inf Comp Sci, 1997, **37** (3): 615~621

Progress of Bioinformatics Database Services. LI Wei-Zhong, WANG Ren-Xiao, LIN Da-Wei, MAO Feng-Lou, HAN Yu-Zhen, LAI Lu-Hua (*Institute of Physical Chemistry, Peking University, Beijing* 100871, China).

Abstract Bioinformatics is one of most active fields in life science. In recent years, various bioinformatics databases have appeared. The size of the database has grown explosively, and the structure of database has been more complex. Now most databases are severed through the internet. The progress in algorithm and software, integration of database and server-client structure make bioinformatics the powerful tool in biology, medicine and agriculture. In 1996 the first network-based bioinformatics server in China was established in Institute of Physical Chemistry, Peking University. Via the Internet, more than 70 000 scientist from all over the world have been served by the server.

Key words bioinformatics, database, software

CAAT 区/增强子结合蛋白 (C/EBP) 的结构与功能

杨根焰 张永莲

(中国科学院上海生物化学研究所, 分子生物学国家重点实验室, 上海 200031)

摘要 C/EBPs 是一组耐热的转录调控因子. 其作用范围广泛, 既参与正常的生理代谢过程, 又与多种疾病的发生和发展相关; 其作用方式多样, 对转录的调控既有正效应又有负作用. C/EBPs 的这种功能多样性是与其结构的特征性相联系的, 它们属于 bZIP 蛋白家族. 自身或与其他异构体形成蕴含着不同调控信息的同源或异源二聚体, 并且能与多种蛋白质因子协同作用, 决定 C/EBPs 发挥作用的方式和细胞特异性.

关键词 C/EBPs, 转录调控因子, bZIP 蛋白

学科分类号 Q71

C/EBP 蛋白 (CAAT/enhancer binding proteins) 因其能与启动子的 CCAAT 区及多种病毒增强子相结合故名. 最早的 C/EBP 蛋白即 C/EBPα 是 1987 年从大鼠肝中分离得到的, 到目前为止报道的 C/EBPs 有六类, 分别为 C/EBPα, C/EBPβ, C/EBPδ, C/EBPγ, GADD153 (CHOP), C/EBPε 以及一个表达还未得到鉴定的基因 CRP1. C/EBPs 具有多种多样的功能, 它们有的结合在同一元件上协同作用, 有的在一定的组织或细胞中发挥其特异的作用; 既有正效应亦有负效应, 与其他蛋白质因子一起组成复杂精细的调控网络, 在细胞增殖、分化、信号传导、肿瘤发生以及机体的免疫、应激反应、能量代谢、血液生成等方面发挥重要作用. C/EBPs 的这些功能是与其特定的结构相联系的 (图 1). C/EBPs 发挥反式调控作用有三个必需的功能域[1]: a. 稳定功能域 (SR). 位于 C/EBPs 的 N 端, 能起到稳定 C/EBP 蛋白结构的作用. b. DNA 结合功能域 (DBD). 位于 C/EBP 的 C 端包括亮氨酸拉链区 (LZ) 和碱性氨基酸区 (BR). 两个 C/EBP 异构体的 LZ 区通过 α 螺旋的交互作用形成 C/EBPs 的同源或异源二聚体, 然后成对的 BR 区结合在 DNA 上. 具有这种保守结构的蛋白质又称为 bZIP 蛋白. c. 激活功能域 (AD) 位于 DBD 区和 SR 区之间. 有趣的是, C/EBP 包含两个功能上可相互促进但不相互依赖的激活功能域 AD1 和 AD2, 且 AD1 的激活功能强于 AD2.

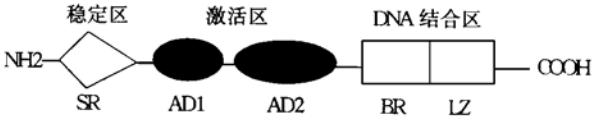


图 1 C/EBPs 的功能域示意图

1 C/EBPα

从整体上来看, C/EBPα 的基本作用是建立和维持分化状态并抑制生长.

1.1 C/EBPα 的结构

C/EBPα 除有图 1 所示的 C/EBPs 的必需功能

收稿日期: 1997-09-14, 修回日期: 1998-02-22