

后基因组时代的生物信息学

陈 铭*

(德国比勒费尔德大学 工程技术学院生物信息学与医学信息学系,比勒费尔德, D-33501)

摘要 随着人类基因组计划的完成,不断积累的巨量的生物学数据和快速发展的信息学技术,给后基因组时代的生物信息学研究带来了新的挑战。该文对后基因组时代的生物信息学研究内容进行了比较全面的描述,分别就其研究对象和研究方向作了区别讨论,分析了生物信息学研究的现状和趋势,比较了国内外的研究发展情况和差距。针对我国在研究中所存在的主要问题,提出了建议并做了展望。

关键词 后基因组;生物信息学

中图分类号:Q343.1 文献标识码:A 文章编号:1672-5565(2004)-02-0007-06

Bioinformatics in the post – genomic era

CHEN Ming

(Department of Bioinformatics / Medical Informatics, Faculty of Technology, Bielefeld University, D-33501 Bielefeld, Germany)

Abstract The successful completion of the human genome project, the exponentially increasing volume of biological data and the advancement of information technology pose new challenges for bioinformatics in the post – genome era. In this paper various research disciplines of bioinformatics are described, recent R&D progresses and current research trends in bioinformatics are discussed and the main research gaps between China and the world are presented. In addition, the author makes a suggestion regarding bioinformatics research, education in China, and outlines a perspective on future development.

Key words post – genomic; Bioinformatics

自从上世纪 90 年代美国能源部和国立卫生院正式启动人类基因组计划以来,现代生物学发生了前所未有的变化和快速发展。15 年来,除人类基因序列测序外,另有上百个物种的基因序列都基本上已经得到。其间,不仅是序列测定,同时进行了其他生物学,特别是分子生物学的研究,获得了大量的实验生物数据。生物信息学,作为一门新兴学科,便应运而生。生物信息学就是利用计算机科学(信息学)的技术手段来研究生物学的的数据,如对生物数据进行获取,存储,传输,计算,分析,模拟,预测等等。目前一般意义上的生物信息学还是局限在基因层次,而广义上的生物信息学是可以研究生物学的任何方面。生命现象是在信息控制下不同层次上的物质、能量与信息的交换,不同层次是指核酸、蛋白质、细胞、器官、个体、群体和生态系统等。这些层次的系

统生物学研究将成为后基因组时代的生物信息学研究和对应用的对象。

目前公开的有 500 多个生物信息学数据库,如大家熟悉的 GenBank、SwissProt、KEGG 等等,涉及核酸序列、基因组、蛋白质序列、蛋白组、蛋白质结构、代谢反应及途径,其他混合型以及其他方方面面的生物数据。每年 "Nucleic Acids Research" 第一期专辑便有大量新的和更新的数据库。如何构建新的生物数据库,如何整合现有的数据库或生物数据,如何开发生物数据软件工具,如何应用于工农医药方面的研究等,是直接摆在我们面前的问题。

发达国家如美国、欧洲、日本在生物信息学研究方面,不容置疑,已经走在国际前沿了。尽管我们国家也参与了人类基因序列 1% 的测定工作,但由于各种原因,生物信息学的起步研究还是落后于其它

收稿日期:2004-02-18;修回日期:2004-03-8

* 作者简介:陈铭(1972-)男,浙江省乐清市人,研究方向:系统生物信息学, e-mail: mchen@techfak.uni-bielefeld.de

万方数据

国家。如何在后基因组时代能快速有效追赶甚至在某些方面超越,都是我们不得不认真考虑并付诸行动的任务。本文将力求全面地讨论后基因组时代的生物信息学研究的方向,对比国内外的差距,寻求较好的发展道路,并对当前国内生物信息学的科研教育提出些建议。

1 后基因时代生物信息学研究内容

虽然生物信息学可以理解为“生物学+信息学(计算机科学及应用)”,但作为一门学科,它有自己的学科体系,而不是简单的叠加。有必要强调的是:生物信息学是一门工程技术学科。必须注意到生物信息学的研究内容与研究对象或客体(应用方面)是不同的概念。

2 生物信息学研究对象

很显然,生物信息学的研究对象是生物数据。当然最“经典”的是分子生物学数据,是基因组技术的产物:DNA序列。后基因组时代将从系统角度研究生命过程的各个层次,走向探索生命过程的每个环节(微观(深入到研究单个分子的结构和运动规律)和宏观(结合宏观生态学,从大的角度来研究生命过程))。着重于“序列、结构、功能、应用”中的“功能和应用”部分。就研究面来说,涉及并参与各生命科学领域的研究。

2.1 分子与细胞生物学

以DNA、RNA、蛋白质为对象,分析编码区和非编码区中信息结构和编码特征以及相应的信息调节与表达规律等。由于生物功能的主要体现者是显型蛋白质及其生理功能,研究蛋白质的修饰加工、转运定位、结构变化、相互作用等活动将推动对基因的功能、表达和调控的理解。对细胞活动,器官、系统、整体活动的调控都很关键。

2.2 生物物理学

生物物理学其实是物理学的一个分支,就象生物信息学是信息学的一个分支一样,研究的是生物的物理形态,涉及生物能学,细胞结构生物物理学,电生理学等等。但这方面的生物数据获取和分析也越来越依赖于计算机的应用了,如模型的建立,光谱、成像数据的分析等等。

2.3 脑和神经科学

脑是自然界中最复杂的物质,其功能是自然界中最复杂的运动形式并随着人类的进化而不断发展

和完善。长期以来,通过神经解剖、神经生理、神经病理和临床医学研究,获得了大量有关脑结构和功能的数据。近年来,神经生物学研究也取得了大量的科研成果,但是这些研究大多是在组织、细胞和分子水平进行的,不能很好地在系统和整体水平上反映人脑活动的规律。随着核磁共振成像和正电子发射断层成像的发展,应用计算机技术,使得我们有可能在系统和整体水平上无创地研究人脑的功能定位、功能区之间的联系以及神经递质和神经受体等。由此产生的神经信息学研究将对了解脑、治疗脑和开发脑产生重大作用。

2.4 医药学

人类基因组计划的目的之一,就是找到人类基因组中的所有基因。如何筛选分离各疾病的致病基因,获得疾病的表型相关的基因信息的工作还刚开始。如何在现有的基因测序的工作平台上,强化生物信息学平台的建设,加快对突发性疫情,公共卫生进行监控和对制病原进行快速有效的分析和解决。而结合生物芯片数据分析,确定药物作用靶,再利用计算机技术进行合理的药物设计,将是新药开发的主要途径。

2.5 农牧渔林业

基因组计划也加快了农业生物功能基因组的研究,加快了转基因动植物育种所需生物信息学研究的步伐。通过比较基因组学、表达分析和功能基因组分析识别重要基因,为培育转基因动植物、改良动植物的质量和数量性状奠定基础。通过分析病虫害,寄生生物的信号受体和转录途径组分,进行农业化合物设计,结合化学信息学方法,鉴定可用于杀虫剂和除草剂的潜在化学成分。可以进行动植物遗传资源研究,保护生物多样性。同样也可以对工业发酵菌进行代谢工程的研究,有目的地控制产品的生产和丰收。

2.6 分子和生态进化

另一个重要的研究对象就是分子和生态进化。通过比较不同生物基因组中各种结构成分的异同,可以大大加深我们对生物进化的认识。从各种基因结构与成分的进化,密码子使用的进化,到进化树的构建,各种理论上和实验上的课题都等待生物信息学家的研究。

3 生物信息学研究方向

后基因组时代,除了继续序列和结构分析外,更

多的研究力量则投入到功能分析,也就是分析研究遗传型到表型的过程。

3.1 序列分析

十几年来,序列方面的分析计算,如序列对准方面的算法,是生物信息学的传统研究领域。经典的算法如记分法和概率统计法:Needleman - Wunsch 算法^[1],Smith - Waterman 算法^[2]和 Hidden Markov Model^[3],Neural Network^[4],还有语言学方法^[5],Z 曲线法^[6]等等。目前有关 DNA,特别是 RNA 序列及针对系统发育推断,大规模序列(全基因组的序列)的更新,更灵敏,效率高的算法也不断推出,必须经常阅读最新期刊才能保持不落伍。

3.2 数据库建设

生物数据库的建设是进行生物信息学研究的基础,尽管目前已有这么多的公共数据库可供使用,如 GenBank。而且它们还同时集成开发了相应得生物分析软件工具,如 NCBI 的 Blast 系列工具,背后都是大量生物信息学的工作。但我们进行专项研究时,往往需要组建新的数据库。建立自己的数据库,就必须分析数据库的储存形式和复杂程度,选择怎么样的数据库,怎么开发信息交流平台,要不要提供相应的分析程序,甚至要不要将各搜索算法硬件化,实行并行计算和先进的内存管理以提高速度等等。如还要考虑到数据库的价格,象 Oracle(www.oracle.com)的大型数据库比较昂贵,MySQL(www.mysql.com)免费但功能可能满足不了要求,目前看来基于 UNIX 开发的共享数据库 PostgreSQL(www.postgresql.org)可能是个选择。甚至考虑用 XML 数据库,也有可能。如果要构建二级数据库,可能还要涉及到其他多个数据库的数据挖掘和整合。

3.3 数据库整合和数据挖掘

生物数据库覆盖面广,分布分散且是异质的。当根据一定的要求要将多个数据库整合在一起提供综合服务,提供数据库的一体化和集成环境,最简单的是用超级链接或进行拷贝重整,但往往简单的连接不合要求,重整涉及数据下载和更新的问题,而且不是真正意义上的“整合”。目前使用较多的是联合数据库系统,它是 IBM 分布式数据库解决方案的重要组成部分,支持用户或应用程序在同一条 SQL 语句中查询不同数据库甚至不同数据库管理系统中的数据。也有直接基于 Internet 技术而进行远程查询而进行文本数据挖掘和重整的。由于生物学的分支学科较多,整合时还必须注意到不同数据库语义学一

致性。其实这已经成为了通过标准查询机制来连接数据库的一大阻碍了。Ontology^[7]技术可能可以解决这一问题。

3.4 结构分析与功能预测

结构分析的研究重点在于研究蛋白质的空间结构。利用分子模拟技术结合计算机图形技术可以更形象、更直观地研究蛋白质等生物大分子的结构,蛋白质的空间结构的更清晰的表述和研究对揭示蛋白质的结构和功能的关系、总结蛋白质结构的规律、预测蛋白质肽链折叠和蛋白质结构等,都是有力的帮助和促进。同时,也可以对已经被测定的生物大分子的三维结构进行显示和编辑操作。分子模型的建立为下一步进行的分子模拟以及了解结构与功能的关系打下了基础。蛋白质结构预测是利用已知的一级、二级序列来构建蛋白质的立体结构模型,对蛋白质进行结构预测需要具体问题具体分析,在不同的已知条件下对于不同的蛋白质采取不同的策略。因为它跟新药开发直接联系,排名前 20 位的世界医药公司都有搞专门从事这方面的研究部门。分子模拟及功能预测还要涉及高性能计算机的硬件平台和处理技术,如网格技术^[8]。

3.5 大规模功能表达谱的分析

生物芯片因为其具有高集成度,高并行处理能力,可自动化分析,可对不同组织来源,不同细胞类型,不同生理状态的基因表达、蛋白质反应进行监测,获得功能表达谱。可进行 DNA,蛋白质的快速检测,药物筛选等。无论是生物芯片还是蛋白质组技术的发展都更强烈地依赖于生物信息学的理论与工具。目前有关表达谱的分析还不很精确,仍需大量的工作来提高对斑点图像处理的能力和系统的分析^[9]。

3.6 代谢网络建模分析

代谢网络涉及生化反应途径,基因调控,信号转导过程(蛋白质间的作用)等等。后基因组时代的研究将研究大规模网络的生命过程,有叫作“网络生物学^[10]”研究。

(1) 预测调控网络

尽管目前已有多个代谢网络途径数据库,有些数据可以直接参考使用,而且这些数据库本身除了手工和自动检索文献以补充数据外,也有开发预测工具的^[11],但是都有局限性和准确性的问题。还需要大量的工作去做从基因组来预测网络,或针对性地去整合某些数据,研究其规律,开发算法模型等。

已有若干研究小组从事“基因组到代谢网络”的预测^[12]。

(2) 网络普遍性分析

建立网络关系后,往往是就简单物种而言,好多人已经对网络的“图论”方面的属性做了分析,如最短距离、连接度等,正试图给出一些重要的结论^[13]。也有分析其最小单元的代谢途径的^[14]。越来越多的人开始开发专门软件工具来自动分析大规模网络系统的物理属性,提供路径导航、模式搜索、图形简化等等分析手段。

(3) 建立模型分析

目前已有若干个比较优秀的代谢网络建模工具,如 Gepasi(www.gepasi.org)、E-cell(www.e-cell.org)、Dbsolve(homepage.ntlworld.com/igor.goryanin)等,它们大都基于代谢控制分析原理,使用常微分方程来求解反应速率。基于标准化数据输出输入考虑,它们已经组成了合作组,共同支持 SBML(www.sbml.org)数据交换。其他形式的建模工具也很多,如用随机方法处理的,因为毕竟确切的动态参数目前还是很难得到。其它如用 Petri net 进行建模的,由于其强大的数学计算功能和明了的示图形式,也越来越多地引起人们的兴趣^[15]。另外,如何自动建立大规模的代谢网络,也是个正在进行中的课题。

与代谢分析直接相关的便是系统生物学^[16]研究。它将是后基因组时代最为突出的研究方向。EMBL(www.embl-heidelberg.de)2006-2015年战略发展目标中已将系统生物学列为三大主要挑战之一。它要求我们看待生命活动过程要用系统的眼光,而不能只盯住一个方面的数据分析而隔离联系。所谓的“虚拟细胞^[17]”模型就是基于系统考虑。

3.7 程序开发

程序开发是生物信息学研究过程中无所不在的任务。这里强调的是生物信息学发展对计算机科学

的进步和发展推动。生物信息学各个领域中的软件工具数目庞大。并行处理、面向对象算法、人工智能等已被应用到最新的程序中。各种人工语言如 SBML,各个共享程序开发小组如 bioperl(www.bioperl.org)等也应运出现。积极参与这些小组的讨论和工作,充分利用公开的资源 and 代码,对提升程序开发能力都很有作用。

3.8 商业化

据 IDC 报告(www.bio-itworld.com),生物信息学技术到 2006 年全球市场需求将达到 380 亿美元,并将在今后 10 年内以每年 24% 的速度递增。由于前期得到的巨量的生物数据以及后基因组时代的数据量以及数据的复杂性,在生物信息技术市场中,存储系统需求最大。其次是服务器,生物数据比较复杂,所以要求服务器的性能相对要高些。其它硬件投入和开发的也包括生物医学工程中仪器设备中的数据储存和分析处理部分。另外,生物芯片预计会成为一个新的产业,大量使用于疾病的分析治疗,药物的设计,代谢工程等领域。而实验室数据信息管理系统(LIMS),也将分享这生物信息时代的市场。

4 生物信息学在国内外的的发展情况

生物信息学在我国的真正起步才近几年,以往更多的是零星涉及和研究,值得高兴的是目前有象中国科学院、中国医学科学院、军事医学科学院、北京大学、清华大学、天津大学、浙江大学、复旦大学、哈尔滨工业大学、东南大学、中山大学、内蒙古大学等等单位在从事研究。但诸多原因,其中包括认知原因,我们已经在这新兴的学科领域落后于人家。可以从表 1 的生物信息学大事记中对比看出我们的贡献和差距。

表 1 国内外生物信息学历史事件对比

年份	国外	国内
1943 年	第一台电子管计算机 ENIAC 诞生	
1953 年	Francis Crick James Watson 和 Maurice Wilkins 发现 DNA 的双螺旋结构	
1955 年	第一个蛋白质序列(牛胰岛素)被测定	
1956 年	“生物学中的信息理论讨论会”于美国田纳西州的 Gatlinburg 召开	
1958 年	由 Hubert P. Yockey 编辑的《生物学中的信息理论讨论会》出版	我国第一台电子管计算机诞生

年份	国外	国内
1965 年		我国人工合成牛胰岛素结晶
1969 年	Kenneth Thompson 开发 UNIX 操作系统	
1970 年	Needleman - Wunsch 序列比准算法(1)	
1972 年	Dennis Ritchie 发明 C 语言	
1975 年	Bill Gates 和 Paul Allen 成立微软	
1981 年	Smith - Waterman 序列比准算法(2)	我国实现酵母内氨酸转移核糖核酸的人工合成
1985 年	Kary Mullis 创立 PCR 技术 ;生物信息学专业期刊(CABIOS)创刊 ;德国生物信息学会议(GCB)举行 ;Bjarne Stroustrup 创建 C++ 语言 ;视窗微软问世	
1986 年	日本核酸序列数据库 DDBJ 诞生 ;蛋白质数据库 SWISS - PROT 建立	我开始实施高技术研究发展的“863 计划”
1987 年	Larry Wall 推出 Perl 语言	
1988 年	美国国家生物技术信息中心(NCBI)成立 ;Pearson 实现 FASTA 程序	
1990 年	国际人类基因组计划(HGP)启动 ;第一届国际电泳、超级计算和人类基因组会议在美国佛罗里达州会议中心举行 ;Altschul 实现 BLAST 程序 ;HTTP 1.0 标准发布	
1993 年	欧洲生物信息学研究所(EBI)获准成立 ;第一届 ISMB 国际会议美国国家医学图书馆(NLM)举行	中国开始参与人类基因组计划
1994 年	Marc Wilkins 提出蛋白质组(Proteome)的概念 ;James Gosling 推出 JAVA 语言	
1995 年	日本信息生物学中心(CIB)成立	
1996 年	W3C 推出 XML 工作草案 ;Affymetrix 生产商用 DNA 芯片	北京大学蛋白质工程和植物遗传学工程国家实验室加入欧洲分子生物学网络(EMBNET)
1997 年	大肠杆菌基因组完成	北京大学生物信息学中心(CBI)成立 ;中国科学院召开“DNA 芯片的现状与未来”和“生物信息学”香山会议
1998 年	亚太生物信息学网络(APBioNet)成立 ;瑞士生物信息学研究所(SIB)成立 ;美国 Celera 遗传公司成立 ;线虫基因组完成 ;CABIOS 期刊更名为 Bioinformatics 中国人类基因组研究北方中心(北京)和南方中心(上海)成立	
1999 年	人类 22 号染色体序列完成	中国获准加入人类基因组计划 ,成为第六个国际人类基因组计划参与国
2000 年	德、日等国科学家宣布基本完成人体第 21 对染色体的测序工作	中国科学院上海生命科学研究院生物信息中心(SIBI)成立
2001 年	美、日、德、法、英、中 6 国科学家和美国 Celera 公司联合公布人类基因组图谱及初步分析结果	首届全国生物信息学会议(CCB)举行 ;中国完成籼稻基因组工作框架图
2002 年	老鼠基因组完成	
2003 年	HGP 完成	

表 2 生物信息学文献历年统计

年份	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004*
文章数量	4	10	10	74	82	295	350	711	1019	1423	1352	218

* 统计数据到 2004 年 2 月 15 日

表 2 列出了 PubMed 收录的以关键词“Bioinformatics”检索到的历年发表的文章数。可以看出大量研究文献出现在 21 世纪后。其中我国共有 138 篇占全部 5548 篇的 2.5% ,而美国则以发表 2160 篇占全部的 39% 之多。我国学者在生物信息学领域发表的有影响力研究数据只有不到美国学者发表数量的百

分之六 ,差距依然很大。我国生物信息学的研究面不广 ,进步不快 ,甚至重复工作原地踏步。虽然多了些学校开设了生物信息学专业 ,可是成果不多。而且目前有从事研究的还不是走主流路线。具有前瞻性 ,大家型的科学家很少。一般的教授 ,教师能力有限 ,有些甚至对生物信息学本身知其一不知其二 ,缺

乏正规的训练,很少了解目前的研究重点、热点和今后方向。由于所申请的经费支持力度和持续时间的原因,大多数学者只能选择易于获得研究结果的科研项目,一般缺乏新颖性和创造性。这可能与我国国家处于快速发展阶段的“短平快”思路和环境有关。另一方面还可能是教育体制上的原因,科研项目的分配问题、行政管理的改革问题、传统的不鼓励学生进行批判性、创新性地学习和思考,也可能是问题症结所在。另外,信息时代信息的交流是决定性因素,尤其是 Internet 的畅通,更是不能有人为的限制访问或限制流量这些今后回顾时会成苦涩笑话的举措。目前我国的科研经费真正投入并落到实处的占国家 GDP 的份额还很小。而德国 1999 年科研项目基金用于生物信息学启动研究就投入 5 千万马克(www.dfg.de),2001 年资助 6 个生物信息学研究中心又投入了 1 亿马克(www.bmbf.de)。目前德国近 50 所大学有生物信息学研究中心和小组,另有至少 50 个生物信息学公司的 1700 名从业人员从事生物信息学研究和产品开发。

5 展望

生物信息学依然是门新兴的工程技术学科,对刚起步的我们来说依然充满机会和挑战,后基因时代已基本给我国的生物信息学发展提供很大的舞台。生物信息学首先是门信息学,所以我们必须端正一些可能的认识错误,必须呼吁引导更多的计算机、数学、物理人才加入到其中的研究。如统计学、概率论、组合数学(尤其是图论)、拓扑学、运筹学、函数论、信息论、计算数学、群论、人工智能,都是已经在生物信息研究中发挥了巨大作用。尽快缩小我国在计算机信息学的核心技术、巨型计算机的应用,以及互联网核心技术方面与世界领先国家的差距。

我国是一个生物信息资源最为丰富的国家之一,这为我们的生物产业奠定了发展的基础。我们应当结合自己的资源条件,建设自有的数据库,如已进行的杂交水稻数据库等,在平等的基础上与国外共享生物信息资源。在考虑社会效应的同时,不断加强基础性研究。如我们的蛋白质组学有优势,可以继续投入生物信息学研究力量,构建完整而强大的科研平台。后基因时代的生物信息学研究将于生命科学和计算机科学一道发展,它将对生物学、医学、药学、农业、环境、信息技术和新材料研究都将起着深刻的作用。生物信息学的进一步深入研究和广

泛应用将为这些领域带来根本性变革。

6 致谢

作者非常感谢徐德昌教授对本文的帮助。

参考文献(References):

- [1] Needleman S.B & Wunsch C.D. Needleman - Wunsch Algorithm for Sequence Similarity Searches[J]. *Mol. Biol.* 1970, *48*: 443 - 453.
- [2] Smith T.F, Watanabe M.S. Identification of common molecular subsequences[J]. *J. Mol. Biol.* 1981, *147*, 195 - 197.
- [3] Koski T. Hidden Markov Models for Bioinformatics[M]. Dordrecht Kluwer Academic Publishers, 2001.
- [4] Wu C.H., McLarty J.M. Neural Networks and Genome Informatics [J]. *Methods in Computational Biology and Biochemistry*, 2000 Volume 1, Series Editor A. K. Konopka, Elsevier Science.
- [5] Dong S., Searls D.B. Gene structure prediction by linguistic methods [J]. *Genomics*, 1994, *23*(3), 540 - 51.
- [6] Zhang R., Zhang C.T. Z Curves, an Intuitive Tool for Visualizing and Analyzing DNA sequences[J]. *J. Biomol. Struct. Dynamics*, 1994, *11*: 767 - 782.
- [7] Stevens R., Goble C.A., Bechhofer S. Ontology - based knowledge representation for bioinformatics[J]. *Briefings in Bioinformatics*, 2000, *1*(4): 398 - 414.
- [8] Chien A., Foster I., Goddette D. Grid technologies empowering drug discovery[J]. *Drug Discov Today*, 2002, *7*(20 Suppl): S176 - S180.
- [9] Hariharan R. The analysis of microarray data[J]. *Pharmacogenomics*, 2003, *4*(4): 477 - 97.
- [10] Barabási A-L., Oltvai Z.N. Network Biology: Understanding The Cell's Functional Organization[J]. *Nature Reviews Genetics*, 2004, *5*: 101 - 113.
- [11] Karp P.D., Paley S., Romero P. The pathway tools software[J]. *Bioinformatics*, 2002, *18*(suppl. 1): S1 - S8.
- [12] Selkov E., Maltsev N., Olsen G. A reconstruction of the metabolism of *Methanococcus jannaschii* from sequence data[J]. *Gene*, 1997, *197*: GC11 - 26.
- [12] Wagner A., Fell D.A. The small world inside large metabolic networks[J]. *Proc R Soc Lond B Biol Sci*, 2001, *268*(1478): 1803 - 10.
- [13] Wagner A., Fell D.A. The small world inside large metabolic networks[J]. *Proc R Soc Lond B Biol Sci*, 2001, *268*(1478): 1803 - 10.
- [14] Schuster S., Fell D.A., Dandekar T. A general definition of metabolic pathways useful for systematic organization and analysis of complex metabolic networks[J]. *Nat Biotechnol*, 2000, *18*(3): 326 - 32.
- [15] Chen M., Hofstaedt R. Quantitative Petri net model of gene regulated metabolic networks in the cell[J]. *In Silico Biol*, 2003, *3*(3): 347 - 65.
- [16] Kitano H. *Foundations of Systems Biology*[M], Cambridge: MIT Press, MA.
- [17] Loew L.M., Schaff J.C. The Virtual Cell: a software environment for computational cell biology[J]. *Trends Biotechnol*, 2001, *19*(10): 401 - 6.