

· 中药现代化论坛 ·

生物信息学发展与中草药研究

许忠能^{*}

(暨南大学生命科学技术学院, 广东 广州 510632)

摘要: 立足于利用其他学科的思路与工具促进中草药研究, 根据生物信息学及相关领域近几年的研究, 就 DNA 测序、基因预测、亲缘及变异与功能、蛋白质结构预测、生物芯片技术、基因多样性、生物安全、活性成分的药靶和预测活性成分结构、代谢过程的模拟、能量中草药药理等方面展开阐述。描述了中草药资源与药理研究的一些可能趋势, 其中药靶与活性成分结构预测、代谢模拟和药理的能观点是中草药研究尤其值得探索的新内容。

关键词: 生物信息学; 中草药; 资源; 药理

中图分类号: R28 文献标识码: A 文章编号: 0253-2670(2003)06-0481-06

Relationship between development of bioinformatics and study of Chinese herbal medicine

XU Zhong-neng

(College of Life Science and Technology, Jinan University, Guangzhou 510632, China)

Abstract Based on the bioinformatics and some frontier science for recent years, the new ideas and methods were used to improve the study of Chinese herbal medicine. Advances in DNA sequencing, prediction of gene, evolution and mutation, prediction of protein structure, biological chip technique, drug target and prediction of drug structure, simulation of metabolism, and the distribution of energy in pharmacological process showed the prospect of the study of Chinese medicinal plants. Some new areas about sources and pharmacology of Chinese herbal medicine are coming into being, especially the relationship between drug targets and drug structure, between simulation of metabolism and pharmacology, and between energy and pharmacology.

Key words bioinformatics; Chinese herbal medicine; resources; pharmacology

利用中草药实质上是极广泛地利用自然界的基因资源, 而探求其分子生物学本质是一项基本任务。同时中草药药效成分确定及多种成分间的网络联系等环节仍未清晰, 利用基因与基因组学的成果可能有助于寻找新的出路。生物信息学是一门综合获取、组织、处理生物大分子及与这相关过程的信息的学科, 其运用了应用数学、计算机科学和生物学的手段去解释大量资料中的生物学、医药学意义^[1]。目前, 多个国家和组织正维护有关网站提供生物学资料、生物信息学分析工具, 这都有利于生物信息学向相关的研究领域渗透^[2]。生物信息学的发展极可能加快中草药研究的进程。

1 DNA测序

DNA测序过程由始至终贯穿着生物信息学方法。现在比较成熟的自动化 DNA序列测定的工作

基础是一个高分辨率的电泳过程。其中双脱氧(Sanger)法和化学裂解(Maxam-Gilbert)法广泛应用。双脱氧法利用 DNA酶合成单链 DNA模板的配对拷贝; 化学裂解法是利用特定化学试剂与相应碱基发生专一反应。另外, 显色技术的发展也为核酸的自动化测序提供方便^[3]。对于上亿个甚至更多的碱基序列, 很难一次就把它测完, 因为目前的 DNA自动测序仪只能直接测出 300~500个碱基长度的结果。一般要将大的 DNA序列打断成 3 kb左右的小片断才能测定。于是大规模 DNA序列或者全基因组的测定就必须选择如何分片断的测序策略。而现在用得最多的测序策略主要有两种: 人工染色体文库方法和全基因组鸟枪法^[4,5]。拼接是将打碎的已测出序列的片段按顺序连接起来的工序。在大量片断中寻找出首尾重叠的部分时须依靠较复杂的数学算

^{*} 收稿日期: 2002-01-14

作者简介: 许忠能(1973-), 男, 理学硕士, 讲师, 研究水生生物及动物生理生态, 已发表论文 19篇。

Tel (020) 85220239 E-mail txuzn@jnu.edu.cn

法才能快捷地解决处理时间问题,并评价测定中的错误^[6,7]。而不同的测试中心甚至不同的个人都可能用不同的方法。常见中草药中一些品种已有某段编码蛋白质的基因被完整测序(表 1)。

表 1 某段编码蛋白质的基因已被完整测序的常见中草药

Table 1 Common Chinese herbal medicine with at least one gene coded protein sequenced completely

品 种	基 因 名	序列长度 /bp
党参 <i>Codonopsis pilosula</i>	atpB	1 497
三白草 <i>Saururus chinensis</i>	atpB	1 497
肉豆蔻 <i>Myristica fragrans</i>	atpB	1 497
杜仲 <i>Eucommia ulmoides</i>	atpB	1 469
猕猴桃 <i>Actinidia chinensis</i>	atpB	1 471
牡丹皮 <i>Paeonia suffruticosa</i>	rbcl	1 402
地黄 <i>Rehmannia glutinosa</i>	rbcl	1 329
山茱萸 <i>Comus officinalis</i>	rbcl	1 398
苦参 <i>Sophora flavescens</i>	rbcl	1 368
知母 <i>Anemarrhena asphodeloides</i>	rbcl	1 408
槐花 <i>Sophora japonica</i>	rbcl	1 368
槐角 <i>Sophora japonica</i>	rbcl	1 368
马钱子 <i>Strychnos nux-vomica</i>	mat K	1 536
麦冬 <i>Ophiopogon japonicus</i>	mat K	2 570
苍术 <i>Atractylodes lancea</i>	mat K	2 630
人参 <i>Panax ginseng</i>	gag 2	1 080
天麻 <i>Gastrodia elata</i>	gastrodiain-MGM	516
丝瓜络 <i>Luffa cylindrica</i>	Rnase LC1	809
关苍术 <i>Atractylodes japonica</i>	ndh P	2 250
芥子 <i>Sinapis alba</i>	Saap1	1 023
青蒿 <i>Artemisia annua</i>	c ASC125	1 843
莱菔子 <i>Raphanus sativus</i>	cox I	4 815
莼菜叶 <i>Hyoscyamus niger</i>	HR7	2 651
银杏 <i>Ginkgo biloba</i>	Legumins	2 747
棉籽 <i>Gossypium herbaceum</i>	f ad2	1 289
紫苏 <i>Perilla frutescens</i>	g PFLC	5 356
颠茄草 <i>Atropa belladonna</i>	mt 2b	544

时间截至 2001年 12月 22日的 EM BL数据库上资料。
Data from EM BL databank by Dec. 22nd, 2001.

2 基因预测

目前认为,基因是决定生物遗传性状的单位,中草药的生长与药性同样在遗传上受基因控制。有些生物活性物质含量极微以至于现有的植物化学方法不能检出,但它们却有较明显的药理效应,合理的解释或许要借助生物工程技术与计算生物学的手段。因此,在大规模 DNA测序完成后的第一件重要的工作就是进行寻找基因。以现在所能利用的手段,若通过实验方法在数以亿计的碱基序列中把全部或大多数基因一一鉴定出来,将消耗大量的时间、人力与物力。运用数学与计算机发展的基因预测方法在一定程度上为解决这一问题提供重要的信息。然而,由于生物实验资料不足以涵盖所有基因产生的情况以及基因结构、内部排列的多样性,致使出现多种基因

预测方法,不同的方法结果差异可能很大,又或者特定的方法只对特定的种类的基因预测有效。

一些生物的有关基因已被测定并存入基因库,其中表达序列标签(expressed sequence tags, EST)数据库是收集实验真正测出的基因的序列数据库^[8]。因此,与该库中有关序列的比较就可一定程度上确定某段 DNA序列是否为基因区。而很多计算识别基因的方法着重于处理编码区与非编码区的差异的一些统计性规律,而此识别过程中运用到很多应用数学理论的成果,如马尔可夫链、隐马尔可夫链、贝叶斯方程、非线性分维、多变量分析方法、神经网络法等^[9,10]。识别启动子、内含子剪接位点、起始点、终止点、多聚 A位点、上游增强子和下游增强子等基因结构功能性位点,是基因识别又一依据^[11]。通过特定位点的比对而得出该位点的相对保守序列,通过识别这些序列预测基因。上述方法都会在处理具体的序列时有不同程度的缺陷,有时将这几类方法一起运用会获得更高的准确度。

3 亲缘、变异与功能

中草药种类与功能鉴定的现状及重要方法是以形态学、组织学和植物化学的方法对这些植物类药材进行品种鉴别和品质评价,其依据是传统的中医药学理论与实践及活性成分的检出^[12]。但毫无疑问,这几种方法注重的是中草药商用或医用价值的静态评价,对造成差别的实质原因、改良品质的有效方法等一系列问题不能提供较满意的解释。植物药物的生物活性成分的稳定性受自身生物学条件限制,而不同的产地、不同的气候季节及不同的营养元素来源对生物的生理及遗传物质结构有潜在的或已表现出的影响,这都可能使同一种中草药生物活性成分有波动。另一方面,不同的种类可含有相同的药理成分,探索这类生理依赖性产物的生物学背景是回答问题的关键步骤之一。因此,在中草药种类与功能区别鉴定的进一步研究中,深入生物遗传学、生理学本质的生物大分子检测分析及基于此而进行的生物信息学研究的介入将会产生极有意义的影响。

不同的中草药品种之间的差异,可以最终理解为核酸序列的差异、核酸序列表达的差异、蛋白质序列的差异及蛋白质结构的差异。亲缘关系分析是研究核酸序列与蛋白质序列的系统发育问题。在具体分析时,经常会选择某段在核内的核酸序列、质体内的核酸序列、线粒体内的核酸序列,通过多序列比对操作进行多个生物种类的相关序列的同源性分析,查明这些种类的亲缘关系及变化程度,有必要的話

根据此类数据构建进化树^[2]。

生物种属和功能之间的差异,并不一定是单个基因间的差异,而可能是有别于基因组如何安排编码区和非编码区^[13]。完整的基因组序列的比较研究是解决问题的重要途径。同样,在中草药品种与功能研究中,比较基因组学的成果将会带来帮助。另外,在占全基因组很大份额的基因间区域的 DNA,即所谓的“垃圾”DNA,是目前基因组学研究中的一大难点,一旦此难题有所突破,可能将带出中草药品种与功能研究的新方向。

单核苷酸多态性(single nucleotide polymorphisms, SNP)是DNA差异的最普遍形式。SNP主要指由基因组核苷酸水平上的变异引起的DNA序列多态性,包括单碱基的转换、颠换、插入和缺失。人类SNP位点研究比较成熟,而药用植物的SNP库目前较小^[14]。SNP作为一种新的遗传学研究工具,寻找药用植物与活性物质有关的及对环境因子敏感的DNA变异位点,将对整个药用植物种质资源研究及应用带来较大的影响。人类SNP在现代医学上可应用于寻找疾病相关基因,而在中草药学上的研究将相当重要。因为SNP是与人类的各种遗传特征直接相关,开发SNP意味着对各种人类群体以及细分到每个人进行核苷酸水平上的鉴定和区别,并通过构建人类基因组SNP图谱,进行全基因组的信息分析,搜寻对环境因素、药物因素敏感的DNA变异位点,查明同种症状的不同个体对同种药物的反应差异。这些都有利于在中草药中开发适合不同人群及个人的个体化药物,充分发挥中草药多种活性成分的综合价值,提高其治疗效果。

4 蛋白质结构预测

药用植物的各项生理活动及多种分类性状都直接或间接与蛋白质相关,蛋白质行使功能与其当时所处的结构状况有密切联系,因此,如何获得蛋白质的结构是蛋白质研究的重要课题。

由于二级结构是理解三维结构的基础之一,大量的蛋白质结构预测论文涉及这项内容。众学者提出的二级结构预测方法大致可分为以下几类:立体化学方法、残基统计法、模式识别法和同源比对综合法^[2, 15, 16]。蛋白质三维结构预测指依靠氨基酸序列来确定蛋白质的空间结构,这是蛋白质结构预测的最终目标。而二级结构的预测准确率最高也不过80%,无疑三维结构的预测受的限制就更大。蛋白质三维结构预测大致分作4类:由一级结构直接预测、二级结构组装、同源预测、结构类识别^[2, 17, 18]。

虽然二级结构、三级结构的预测目前都遇到不同程度的困难,但这些方法在迅速发展的实验技术、日益完善的理论推动下,一步一步地改进,在不少领域中,这些预测结果已符合人们的要求。现代实验技术广泛应用于测定药用植物有关成分,测定蛋白质结构的技术设备不断更新换代,测定蛋白质所投入的资金也不断增加,但在未真正有达到要求的仪器方法出现以前,蛋白质结构预测仍值得介入中草药研究中去。

5 生物芯片技术

中草药的深入研究将不可避免地要利用生物芯片,因为生物芯片带来的生物信息蕴藏着生物学结构与功能的内在联系。虽然出现不到10年,就已引起来源极为广泛的人们的注意力,相信不久后研究中草药的不少反应可在芯片上完成。

植物基因的定量表达是植物性状的物质基础,同时也是植物对不同环境的反应之一。cDNA微阵列芯片研究拟南芥菜 *Arabidopsis thaliana* (L.) Heynh. 有关基因定量表达时显示,拟南芥菜用病原真菌 *Alternaria brassicicola* 水杨酸、茉莉酮酸甲酯(methyl jasmonate)或乙烯处理后,在2375个选定的基因中,有705个基因与对照组存在基因表达差异^[19]。应用生物芯片检测植物基因的定量表达可用于药用植物不同环境因子下其种质的差异分析。

生物芯片也可应用于遗传育种,通过探针与目标DNA序列杂交而鉴定出优良品种。为筛选出燕麦与玉米的特定杂交品种,用玉米DNA中的特异序列作为探针,可以在一群混杂的杂交后代中有选择挑出含目标玉米DNA序列的后代^[20]。在有确定的目标核酸序列时,药用植物优良品种的筛选也可将大部分工作用生物芯片来实现。

6 基因多样性

在药物开发史上,在植物极具多样性的地区,会增加可收集到的植物种数,因此更有可能找到具有可利用的药效分子的植物。物种丰度的上升,意味着基因数量增加的可能性变大;个体的增加也会提高SNP的库存。因此在多样性高的条件下,候选基因库更具包容性,理论上提高了有用基因在基因库中的绝对含量。

影响生物多样性的因素除了环境压力外,生物本来的遗传物质基础也是关键一环。生物分类学与物种进化及分子生物学的证据表明,原始的核酸与蛋白质序列随着时间的推移,其中的残基位点不断发生突变而形成现在的基因多样性进而生物多样性。

性;并且基因的进化除了由祖代传给后代的纵向进化外,还有基因的移位。基因在 DNA 链上的不断增加重复数等的横向进化^[21]。因此,要利用、管理、预测生物多样性,必须清楚基因多样性的变化动态。这时,生物信息学又成了重要的研究工具。

7 转基因植物的安全问题

转基因技术为药用植物的高质量培养与生物活性物质的生产、提取提供帮助,尽管技术并不成熟。但是,应该提前认识到转基因植物的安全问题。

当外源基因插入受体植物基因组中,插入的位置是随机的,其拷贝数也不确定^[22]。这时,令人担心的并不是对所转基因的表达功能或该转基因植物是否存活,而是导入的基因是否会与受体植物中的某些基因共同作用产生意想不到的负面影响。另外,当插入的 DNA 片段把受体植物的某个或某些基因隔开引起这部分基因的失活,这时要考虑这类失活是否引起某些生理反应中产生有毒的代谢产物。同时,转基因植物可能由于获得了特定的基因使生命力较强,在其数量超过被利用量时,流入环境中将成为杂草,并通过基因飘移把这些遗传性状传给其他植物。由于杂草具有强大的生存竞争力,常造成种植业的巨大损失^[23]。故有必要在转基因实行前先评价转基因生物新的基因组与杂草基因组间某些功能的差别。这也可交给生物信息学完成。所以,生物信息学的发展可为药用植物转基因安全性提供越来越重要的预测。

8 确定药靶及预测活性成分结构

中草药目前研究出现的其中一个问题是虽然知道了某种中药或某个中药方剂有效,却无法确定活性成分。一则是经化学手段分离后,得出的是一堆数目极其庞大的化学物质;另外,这些化学物质中不少成分结构未知。药靶是联系药物与药理作用的关键部位,中药药理可能是多种成分多个药靶的混合效应。通过对生物与化学有关资源信息的综合分析,首先初步确定药靶,再结合基于受体结构的药物分子设计(sturcture-based drug design, SBDD)领域的成果,根据被结合药靶预测配体结构的方法将会加快搜索中草药活性成分速度,并且更加明了发生药效的原理,有关中草药在这方面的研究鲜有报道,相信这是一个有前景的课题。

某些中草药的部分活性成分的药靶位点已测出(表 2)^[24]。但大多数草药的活性成分药靶仍未确定,进一步的工作可能是运用特定药理过程中一系列生物学反应与特异性受体、酶等关系估计药靶种类和

部位;另外可利用生物芯片测定药物对基因表达的影响效果,进行一些生物信息计算,初步确定药物靶点。然后根据条件用这些确定为药物靶点的物质与药物提取物做结合实验,进一步确定药靶。后面的实验可以在生物芯片上进行。

表 2 一些中草药的部分药靶

Table 2 Drug targets of some Chinese herbal medicines	
药 物	药 靶
葛根 <i>Pueraria lobata</i>	苯二氮䓬受体
黄连 <i>Coptis chinensis</i>	肾上腺素受体 α 型
麻黄 <i>Ephedra sinica</i>	肾上腺素受体 α、β 型
延胡索 <i>Corydalis yanhusuo</i>	多巴胺受体
莨菪 <i>Hyoscyamns niger</i>	毒蕈碱样受体
马钱子 <i>Strychnos nux-vomica</i>	甘氨酸受体
山茶 <i>Canellia sinensis</i>	腺苷受体
银杏 <i>Ginkgo biloba</i>	血小板活化因子受体

在药靶的结构未知时,可利用药理试验中的构效关系确定活性物质的药效团,即对若干种已知的不同原子、化学基团的化学物质进行药理学试验以获得构效关系,待测中草药再通过同样的药理试验,其结果对应于所得的构效关系,从而确定符合其活性的一些原子种类、数目、空间位置的参数;当药靶结构已知时,可根据药靶特定位点的空间化学性质确定药效团的几何及化学上的结构特征^[25]。当药效团确定后,活性物质的结构分析的线索比较明显,这时可以药效团为共同特征寻找活性相似物。蛋白质、核酸数据库可用于中草药活性物质药效团的搜索,还有另外一些化学数据库(表 3),大多是免费的。一般按药效团定义、数据库搜索和优化筛选几个步骤,最后得到一批符合条件的化学物质。假如这种活性物质并非新的化合物,很可能就是所搜寻到的活性相似物中的一种或若干种。

表 3 可用于有特定三维结构化合物搜索的一些化学数据库与网站

Table 3 Chemical databank website for searching compounds with special 3D structure	
数 据 库	网 址
Chem Finder 网络服务器	http://chemfinder.camsoft.com
剑桥结构数据库(CSD)	http://www.csd.cam.ac.uk
美国标准技术研究所(NIST)数据库	http://www.nist.gov/srd/online.htm
药用植物和有毒植物数据库(medicinal and poisonous plant databases)	http://www.wam.umd.edu/~mct/Plants/index.html

9 调控代谢过程的模拟

中草药自身的生理过程及其活性物质的药理过程与细胞内的代谢途径和细胞调控直接有关。蛋白质可以体现细胞活动对内外刺激因子的响应,并同时实现对细胞系统的调节,故在体内大部分调控过程都可以认为是以基因表达为基础。在基因表达起

关键作用的启动子的多个组成结构已被分析出来,各结构之间的关系也不断明了,在细菌与酵母中的研究显示,即使激动剂与 DNA 结合从而激活启动子这一环节中就牵涉到许多基因的表达^[26]。活体细胞调控过程实际是环环紧扣的各种基因表达的网络过程,直至细胞被分裂成不同的主体,部分网络连线才告终结。一些生化反应、代谢途径、生物催化与生物降解、细胞信号网络等资料正在不断丰富,不断被汇总。

细胞代谢调控资料的其中一个用途是作为生物在分子水平上的生理过程模拟。日本科学家开发的 E-CELL 模拟仿真软件,能模拟细胞调控及细胞繁殖 (<http://www.e-cell.org>); 美国科学家设计的 Virtual Cell 可以模拟分子在活体细胞内的分布过程及反应过程,并利用电影的画面生动地将微观现象放大, (<http://www.nrcam.uchc.edu>)^[27]。不同的成分对红细胞的代谢试验可以在计算机内完成,较为详细的胞内网络过程能显示不同环境下红细胞内多种成分的动力学过程及调节因子^[28]。细胞在体内与体外所处的环境有极大的差别,故代谢不同,MEC 程度可模拟细胞在这两种情况下的代谢过程,这就能在某种程度上比较药物体外试验跟活体试验之间的差别 (E-mail: mendes@vt.edu)^[29]。在中草药研究中,通过这些模拟模型,设计不同的环境条件或营养条件,看看药用生物体内成分的变化,特别是可以用这些模拟细胞或模拟生物做中药药理试验,观察特定中草药成分对细胞或生物体分子网络调控过程的影响,因为不少中草药的药理过程很不清楚,用简单的细胞模型分析较少作用元素中的复杂过程变化对整体把握真实过程有帮助。把这些模拟结果作为真实实验室试验设计的依据,节约实验成本,提高实验设计水平,可能因此导致许多新的发现。

10 能量与中草药药理

生物信息学在中草药中的运用将会诱发新的研究课题。用能量去解释中草药药理是一个值得深入探讨的领域。能量除了实测外,更多可计算出来,依据实验、临床数据借助生物信息学的计算工具构建较精确的生命能量系统,以此作为思维平台,形成疾病与治疗的新观点。

能量耗散学说认为,生理过程的规律性与生物构造的稳态依靠消耗能量来维持,体内的各个部分依靠不同水平的能量转化实现对称性结构的存在,当体内能量供应状况剧变时,系统对称性被破坏,而又可能随着能量的涨落,混乱的结构又重新进入新

的稳态。能量的获得、贮存、流动、转化与丧失是一切生命过程的基本特征。在物质形式的研究遇到研究手段的极限时,以能量为基础构建局部的生理过程,在此过程中寻找与之相关的物质,这可能是一个可统一的思路。

能量是病理研究的一个重要的内容。在肿瘤中,由线粒体 DNA 编码的蛋白的表达量会提高,以增强呼吸链的氧化产能效率,适应肿瘤细胞增长的能量需要^[30]。有 Yoshida 肉瘤的小鼠的基础代谢耗能比对照组高^[31]。而带有 Apolipoprotein E 基因变异的个体对食物能量有很大限制,过高的能量会诱发 LDL 胆固醇升高导致心血管疾病^[32]。因此疾病往往可以理解成体内能量流动发生偏差。

而在现代的中药药理中,能量的观点并没有得到体现。在西方医学影响下,传统的中医中药理论在一些研究中主要用生物大分子、细胞器官、细胞、组织及生物体的存亡增减或结构功能改变及其网络关系的物质观点去解释疾病及其治疗药理。有些传统的中药药理的解釋涉及与能量相关的气功和经络,如果仅凭借已知生物分子的形态功能把这些过程生硬地还原成生化途径,将存在极大的障碍。恰恰相反,根据计算,有效的生化、药理途径还原成能量过程是可行的。若有了局部的能量体系,中草药的药效,尤其是多种药效成分的作用,可以理解为对能量的疏导,即调节能量的流向和在各环节中的流动速度等。经过这种还原,在某种程度上更吻合中医药学的传统观点,同时也为生物分子的形态性质及其生化药效途径的研究有效地缩小了范围。生物信息学的预测模型及对生物大分子功能的解释将有效构架能量与中药药理之间的关系。

11 展望

传统的中草药观念及研究方式无疑在目前及以后都是中草药研究的主流之一。同时,从分子生物学的思想出发,越来越多的有关中草药的生药研究、药理学试验、临床报告及各类资料相互的关系所产生的信息,已使一般数据分析方法、技巧及解释理论的利用价值被超越了极限。这时需要借用数学、计算机、网络、物理等方面的成果,合力解决难题。生物信息学这门交叉学科的功能正符合中草药逐渐深入研究的发展需要。所以,无论从以解决问题为核心的角度,还是从寻找学科新的生长点的角度,都有必要考虑用生物信息学的手段去丰富有关中草药的研究思路与研究内容。

当然,生物信息数据库中的数据并非绝对可靠,

同时预测模型也没有太高的准确度,各种软件处理同样的数据时可能会有不同的结果。但由于生物信息学研究大部分是已有实验结果而不需昂贵投入的计算过程,而且可以不断修改,其最后的正确性也是由实验或调查等实践去验证。它与中草药研究的结合意义在于更快、可考虑更多因素地去解释实际问题或设计实验等。不应仅因为其不足而产生学科分界而摒弃。了解中草药资源与药效药理,解决实际问题是中草药研究的任务与本质,强调学科界限反而容易阻碍对中草药本身各种特殊规律的探求。

References

- [1] Luscombe N M, Greenbaum D, Gerstein M. What is bioinformatics² A proposed definition and overview of the field [J]. *Methods Inf Med*, 2001, 40(4): 346-348.
- [2] Baxevanis A D, Ouellette B F F. *Bioinformatics a Practical Guide to the Analysis of Genes and Proteins* (生物信息学—基因和蛋白质分析的实用指南) [M]. Beijing: Tsinghua University Press, 2000.
- [3] Metzker M L, Lu J, Gibbs R A. Electrophoretically uniform fluorescent dyes for automated DNA sequencing [J]. *Science*, 1996, 271: 1420-1422.
- [4] International Human Genome Sequencing Consortium. Initial sequencing and analysis of the human genome [J]. *Nature*, 2001, 409: 860-921.
- [5] Weber J L, Myers E W. Human whole-genome shotgun sequencing [J]. *Genome Res*, 1997, 7: 401-409.
- [6] Ewing B, Hillier L, Wendl M C, *et al.* Base-calling of automated sequencer traces using phred I. Accuracy assessment [J]. *Genome Res*, 1998, 8: 175-185.
- [7] Ewing B, Green P. Base-calling of automated sequencer traces using phred. II. Error probabilities [J]. *Genome Res*, 1998, 8: 186-194.
- [8] Adams M D, Kelley J M, Gocayne J D, *et al.* Complementary DNA sequencing expressed sequence tags and human project [J]. *Science*, 1991, 252: 1651-1656.
- [9] Burge C, Karlin S. Prediction of complete gene structures in human genomic DNA [J]. *J Mol Biol*, 1997, 268: 78-94.
- [10] Reese M G, Kulp D, Tammana H, *et al.* Gene-gene finding in *Drosophila melanogaster* [J]. *Genome Res*, 2000, 10: 529-538.
- [11] Frech K, Quandt K, Wemer T. Finding protein-binding sites in DNA sequences: the next generation [J]. *Trends Biochem Sci*, 1997, 22(3): 103-104.
- [12] Zhao T Z, Liao X T, Qin H L. Identification of traditional Chinese herbal medicine by ¹H NMR fingerprints [J]. *Chin Tradit Herb Drugs* (中草药), 2000, 31(11): 868-870.
- [13] Clark M. *Comparative Genomics* [M]. Norwell: Kluwer Academic Publisher, 2000.
- [14] The International SNP Map Working Group. A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms [J]. *Nature*, 2001, 409: 928-933.
- [15] Rost B. Review: protein secondary structure prediction continues to rise [J]. *J Struct Biol*, 2001, 134(2-3): 204-218.
- [16] Sermala M, Juhola M, Vihinen M. On preprocessing of protein sequences for neural network prediction of polypeptide type II secondary structures [J]. *Comput Biol Med*, 2001, 31(5): 385-398.
- [17] Yan L F, Sun Z R. *Protein Structure* (蛋白质分子结构) [M]. Beijing: Tsinghua University Press, 1999.
- [18] Fariselli P, Casadio R. Prediction of disulfide connectivity in proteins [J]. *Bioinformatics*, 2001, 17(10): 957-964.
- [19] Schenk P M, Kazan K, Wilson L, *et al.* Coordinated plant defense responses in *Arabidopsis* revealed by microarray analysis [J]. *Proc Natl Acad Sci USA*, 2000, 97(21): 11655-11660.
- [20] Ananiev E V, Riera-Lizarazu O, Rines H W, *et al.* Oat-maize chromosome addition lines: a new system for mapping the maize genome [J]. *Proc Natl Acad Sci USA*, 1997, 94: 3524-3529.
- [21] Luo L F. *Physical Aspects on Life Evolution* (生命进化的物理观) [M]. Shanghai: Shanghai Science and Technology Publishers, 2000.
- [22] Liu Q, Zhu Q Q. *Biosafety* (生命安全) [M]. Beijing: Science Publisher, 2001.
- [23] Rifkin J. *The Biotech Century: Harnessing the Gene and Remaking the World* (生物技术世纪: 用基因重塑世界) [M]. Shanghai: Shanghai Science and Technology Publishers, 2000.
- [24] Liu D L. *The Disturbance and the Restoration of the Equilibrium to Molecular Chinese Medicine and Pharmacology* (分子网络紊乱与调节) [M]. Beijing: Tsinghua University Press, 1999.
- [25] Zhou J J, Wang T. *Molecular Modeling Methods in Drug Design* (药物设计中的分子模型化方法) [M]. Beijing: Science Press, 2001.
- [26] Ptashne M, Gann A. Transcriptional activation by recruitment [J]. *Nature*, 1997, 386: 569-577.
- [27] Normile D. Building working cells “in silico” [J]. *Science*, 1999, 284: 80-81.
- [28] Jamshidi N, Edwards J S, Fahland T, *et al.* Dynamic simulation of the human red blood cell metabolic network [J]. *Bioinformatics*, 2001, 17(3): 286-287.
- [29] Mendes P, Kell D B. MEC (model extender for Gepasi): a program for the modeling of complex, heterogeneous, cellular systems [J]. *Bioinformatics*, 2001, 17(3): 288-289.
- [30] Penta J S, Johnson F M, Wachsman J T, *et al.* Mitochondrial DNA in human malignancy [J]. *Mutat Res*, 2001, 488(2): 119-133.
- [31] Loktinov A, Scollen S, Mckeown N, *et al.* Gene-nutrient interactions: dietary behavior associated with high coronary heart disease risk particularly affects serum LDL cholesterol in apolipoprotein E epsilon 4-carrying free-living individuals [J]. *Br J Nutr*, 2000, 84(6): 885-890.
- [32] Oudart H, Malan A, Maho Y, *et al.* Day-night pattern of energy expenditure and body temperature in cachectic tumor-bearing rats [J]. *Br J Cancer*, 2000, 83(8): 1055-1060.