

·科学论坛·

# 关于医学科学中系统复杂性研究的探讨

朱宗涵

(北京市卫生局,北京 100053)

**[摘要]** 本文回顾了科学和医学研究中处理复杂性问题的新的概念和方法。生物信息学已经在基因组学和蛋白质组学的研究中发挥了重要的作用。而基于非线性关系的许多模型和算法也显示出在生物进化和生态学研究方面的巨大潜力。这些有关复杂性的概念和工具将会越来越广泛地被应用于病理过程和临床实践的综合研究中,如癌症、高血压病、代谢综合症等复杂性疾病。

**[关键词]** 复杂性,生物信息学,医学,癌症,单基因病

对复杂性问题的研究已成为一个热门的话题。“复杂性”这一概念尚无明确的定义,但基本上有一个共识,即把那些数据量大、呈现非线性关系、具有自组织、自适应和涌现特征的问题归类为复杂性问题或“复杂系统”。这类系统很难用传统的工具建模,因而也很难预测。“复杂性”概念的应用,有助于学术界的讨论、交流和合作<sup>[1]</sup>。

## 1 基于还原论的研究思想和方法学对医学科学发展的影响——对现代医学的反思

近年来医学界许多学者越来越重视复杂性问题的研究。这首先是由于医疗服务面临复杂的问题和挑战。例如:近20年来投入科研资源最大,科技成果最多,消耗资源最多,诊疗费用增长最快的几大疾病,如心脑血管病、肿瘤、艾滋病等,其总的发病率和死亡率并没有明显的下降;工业化、全球化的进程加快,新发和再发传染病的危害和威胁却在不断增加,使公共卫生机构应接不暇;随着医药领域高新技术产品的快速增加,医疗费用也在飞快地增长,已经成为群众、社会和政府不可承受的沉重负担;在科学工作者为医学科学技术飞快发展而欢欣鼓舞的同时,群众和政府对于医疗服务的不满意程度却有增无减等等。医学领域必定在系统方面发生了问题。首先,医疗卫生系统始终跟不上社会经济领域的变革,改革似乎总找不到出路,不能适应社会和经济体制的

改革和发展。同时,我们也应该从医学科学研究和医学教学这一引导医学发展的前沿领域的系统本身去找问题。一个突出的问题是现代医学科学长期以来对局部的重视、对整体的忽视;对“还原”的重视、对“整合”的忽视。忽视在“系统”的层面上综合地认识和处理人体健康和疾病的复杂问题。

笛卡尔曾指出:“如果一个问题过于复杂以至于一下子难以解决,可以将此问题分解成一些足够小的问题,分别去解决”。笛卡尔的这一名言标志着现代科学还原论思想和方法学的起源。从此,还原论的思想和方法学统治了200多年的现代科学。以还原论为基础的科研思想及方法学,对推动科学的发展起到了重要的作用。在医学领域中,经典的解剖学、生理学、病理学的形成,以及现代的生物化学、细胞生物学、分子生物学等的兴起和发展都与此分不开。笛卡尔的思想方法,或称还原论,隐含了一个假设:“一切系统都是可逆的。分割的各个问题的解答之和就会成为整体的最后答案,即系统之解”。然而生命系统并非如此,系统被分解之后,生命本身就不可挽回地丧失了,也就是说,生命系统是不可逆的。整体不等于局部之和。在还原论思想的影响下,医学分科越来越细,针对局部的治疗越来越精细,遗憾的是对整体的忽视也越来越严重。真正意义上的医生越来越少了,更多的医生成为只能处理某一器官的“器官医生”,不懂全局只知局部的“专科”医生。甚至出现了“开刀匠”和“管子工”的嘲讽。

本文于2005年6月2日收到。

## 2 人体及其健康和疾病过程符合复杂系统的基本特点

肿瘤研究的历程给我们很大的启示。当基因研究时代到来的时候,研究人员企求寻找“致癌基因”,从而找到解决癌症的治本方法。经过四分之一世纪的努力,人们终于发现癌症是一种复杂疾病,涉及一系列基因谱的动态的改变。正常细胞的癌变是一个极为复杂的过程,而当癌变细胞形成一个肿瘤的时候,不仅仅是癌变细胞的集合,而且组合了快速成长的血管和其他组织,成为系统上已经区别于原始组织的一个相对独立的“肿瘤器官”,是机体内部一个叛逆的“独立王国”。癌变过程:自我形成足够的生长信号,对生长抑制信号不敏感,细胞凋亡的逃逸,细胞增殖的失控,血管生成的激活,组织的浸润和转移。结论是并非存在单一的致癌基因,肿瘤的形成涉及细胞分化和生长的多系列的基因组和相应的一批蛋白质和多肽的改变,是多步骤多系统的变化<sup>[2]</sup>。

科技人员仍在研究高血压病基因、哮喘病基因、精神分裂症基因等等,但是总的概念已经发生了变化,人们已经把单个致病基因的研究转变到基因谱的研究,转变到生物信号系统的研究。这些疾病被定义为“复杂性疾病”<sup>[3]</sup>。

对单基因病的研究也有了发展,单基因病也并非简单疾病,实质上也是一种复杂性疾病。例如苯丙酮尿症(PKU)是一种遗传代谢病。PKU是由于苯丙氨酸羟化酶的基因发生点突变而导致的一种严重的隐性遗传病。该基因有13个外显子。我们对141个病例和家系进行研究,共发现42种不同的点突变。由此,造成不同的生化表型、临床表型,以及复杂多变的疗效和预后。

回顾历史,医学科学走过了从原始的朴素的整体论为基础的模式,如中医的“阴阳五行”学说,到基于还原论的现代科学研究的漫长的发展过程。20世纪后期,科技的进步,又推动我们向基于系统论的科学研究进军。系统论是在继承还原论的精华的同时,强调系统的内部和系统之间的联系,系统的复杂性动力学特性,以及整体功能和行为的突现。系统复杂性研究已经引起生命科学和医学科学工作者的关注,并正在逐渐渗透到生命科学和医学的研究中。复杂性(complexity)、多样性(diversity)、突现(emerging)、混沌(chaos)、非线性动力学(nonlinear dynamics)和不确定性(uncertainty)等概念和术语已经越来越频繁地出现在生命科学和医学的研究论文中。

近二十年来,有几个研究领域的工作,对生命科学和医学科学中的系统复杂性研究起到了很大的推动作用:生物信息学(Bio-informatics),生物信号系统研究(Biological Signaling Systems),基因组学(Genomics),蛋白质组学(Proteomics),发育生物学(Developmental Biology),脑科学和行为生态学(Behavioural Ecology)<sup>[4-8]</sup>。

人体作为一个复杂系统,具有自组织和自适应的特性。不确定性和混沌状态是人体系统复杂性的一个重要的特性,也是系统自适应性的重要机制。生命过程:秩序和混沌的边缘。大约30年前,生态学者们开始采用非线性动力学解释生物种群数目的变迁。当研究者们开始为有机体相互作用的方式建立非线性模型时,他们也遇到了人们在其他领域遇到的混沌现象:一种有几个简单的非随机方程产生的看似随机的模式。混沌理论意味着将对传统的生物学界公认的生态平衡观点提出挑战。Carl Zimmer在*Science*上发表了一篇文章“Life after chaos”<sup>[9]</sup>,综述了这些有意义的工作。一些医学科学工作者也尝试应用混沌理论研究电生理中的课题,如脑电图、心电图、心律失常、癫痫等。虽然,某些研究拘泥于混沌数学模式的建立和验证,尚未受到医学界广泛的重视,但是,这些学者毕竟做了大量开拓性的工作。类似的研究也在传染病的流行病学领域中进行,取得了初步的成果,开拓了人们的思路。

医学科学中系统复杂性和非线性动力学及混沌理论的研究得到重视是与生物信息学的产生和发展分不开的。随着生物信息学技术的发展,在基因组学、蛋白质组学的研究中,采用了许多大通量的信息处理技术。所谓“海量”信息的产生和处理,也推动了计算机生物学的发展。海量数据的处理必须利用数学模型、算法(algorithm)和计算机学习技术(machine learning)等。在此过程中,非线性方程和混沌现象成为越来越常见的课题。非线性动力学和混沌机制为人体复杂系统的自适应能力提供了可能性。人体不断面临着内部和外部的侵扰。只有当对付潜在的侵扰的策略和机制失效时系统才会发生崩溃。这是人体自组织和自适应性的又一表现。人体复杂系统中表现为三种反应方式:适应,代偿和失代偿。这体现了人体健康的三种状态:健康,亚健康(疾病前期 pre-disease)和疾病。非线性动力学发展起来的概念和技术将会大大地增强我们处理医学中复杂的非线性过程的能力。

### 3 医学和生命科学中的信息概念

信息学的介入,使生命科学和医学中的系统复杂性研究成为可能,成为强有力的技术支持。

现代生命科学和医学的快速发展几乎都离不开生物化学。以至于有人提出生命的本质是新陈代谢,即化学反应。“爱情就是化学反应。”也正是由于化学研究的局限性,使这些研究长期处于对化学反应的描述中,无法对生命现象进行更加深入的探索和抽象,也使我们的研究工作无法从还原论的阴影中走出来。大多数人都不会同意“生命就是化学反应”的说法,但是除了化学反应外我们还能用什么其他方式描述生命活动呢?在思考这一问题的过程中,我们已经发现当今发展最快的科技领域——信息学,正在步步深入到生命科学领域,并且已经看到了两者结合的远大前景。让我们看看信息学为我们做了些什么?将会为我们做什么?如果我们暂时离开一下生物化学,而站在信息学的角度审视生命的过程,我们将会看到一些什么样的不同的景象?

信息技术的应用首先是医学和生命科学发展的需要。各种基因和蛋白质测序技术的完善和推广,使序列分析的需求迅速增加,从而导致计算机分析应用的第一个高潮的出现。计算机分析能力的不断提高,也进一步鼓励和推动各种高通量和组合实验技术的研发和应用,如微阵列技术(芯片技术)、质谱技术等的重大进展。这使得在实验设计、数据处理和结果解释等方面的计算机支持成为基本的要求。用于序列分类、弱相似性探测、区分DNA序列中的编码区和非编码区、分子结构预测、转录后修饰和功能的预测,以及重构进化树的计算工具已经成为生命科学研究的基本组成部分。在数据呈几何级数增长的情况下,生物信息的存储、获取、联网、处理、浏览以及可视化等方面,都对计算机科学的理论、算法和软件的发展提出了迫切的需求。而计算机科学也从生命科学中获得启示,产生了许多新的概念、算法和技术,例如:机器学习技术(The Machine Learning Approach)、隐马氏模型(hidden Markov model)、支持向量机(support vector machines)、置信网络(belief network)、遗传算法、神经网络算法、人工生命等等。对生物数据进行概率建模的真正需要来源于生命系统的复杂性和多样性,这一切来自于漫长的进化过程中生物体在复杂环境下经历的变异<sup>[10]</sup>。有两种策略推动生命科学与信息学的结合。

第一种策略是直接应用信息科学的新技术,处

理和分析生物学和医学中的数据和问题。例如上面已经提到过的生物信息学,就是利用信息学中的许多新的数学模型和算法来处理基因组学和蛋白质组学中的数据和各种复杂问题。在医学领域中已经被广泛使用的CT、核磁共振、超声波测定技术等都是计算机成像和图像识别技术与临床医学相结合的成功例子。这些影像诊断技术也已经被广泛应用于生命科学的研究工作,例如用CT或MRI技术研究大脑的智能活动、语言发育等。在生态研究和流行病学的研究中也尝试采用非线性关系的数学模型和混沌理论等知识和技术处理和分析实验数据和资料。

第二种策略是真正涉及生命活动中的信息研究,例如研究遗传信息的储存、传递、表达、复制、进化、外延等活动的过程,探讨生命过程中信息作用的机制和原理,探讨生命信息的存生形式和动力学规律,探讨生命信息和环境的关系等,从而开创生命科学中生命信息研究的新领域。这使我们有可能也有必要创立一个新的边缘学科,即“生命信息动力学”。

生命存在的基本抽象:生命物质,生命能量,生命信息。全部的物种和全部的生命存在形式都可以抽象为生命信息的集成:信息流。生命延续的基础是生物一代又一代传递着的基因流,以及基因流承载着的生命信息流。生物体不断死亡,永恒的是基因流和信息流。物种繁衍的过程是一个基因流和信息流的不连续的过程,这一不连续过程为进化提供了可能性。复杂生物的有性生殖的机制,更赋予基因和生命信息流内部差异性和复杂性的无限趋向,为物种的进化和适应性注入了更强的活力。生物的进化是生命信息系统对生物差异性的记录和积累,是生物适应性的记录和积累,本质上是对生物与环境相互适应的建模。所有的物种都是在个体不断的死亡和新生的更替中维持着基因流和信息流的延续。小至病毒和细菌,大至树木和禽兽。基因流也是一个开放的复杂巨系统。在这一巨系统中存在无数的子系统。导致遗传病的许多致病基因构成的基因流,形成巨系统中的子系统,呈现着各自复杂的发病特点和规律。病毒和细菌引起传染病的流行的过程也可以抽象为致病微生物的基因流和信息流在人群中的传布。病毒和细菌的寿命仅为数小时,但是,无数个病毒在死亡和新生的更替过程中将基因和信息流散布到人群中。疾病的发生和流行实质上是基因流在人群中的一种表型。复杂物种的生命信息系统是基因的遗传信息和非基因的后天信息的综合,

后者即是以社会传统和文化的形式存在的生命信息,其载体是广义的“语言”。

#### 4 生命信息存在形式和动力学特征

生命信息的载体,无论是遗传信息的基因,还是社会信息的“语言”都具有一个共同的特征,即简单符号的有限排列,一种不连续的非周期性的序列。这一特征满足了信息具有差异性和差异无限趋向性的基本特征。基因是4个碱基的序列,蛋白质是20个氨基酸的序列,英语是26个字母的序列,汉语可以简化为5笔序列,西方音乐是12个音符的序列(汉乐为5个音符),绘画可以抽象为若干元色的组合等等。所有的信息的载体,无论是基因语言、文字语言、语音语言、艺术语言(音乐、图像等)等,都能简化为“0、1”两个字符的序列,即计算机语言。并都能建立基于二进制和概率计算的数学模型,进行运算和模拟。所有的生命物质都具有这种字符序列的特征:以碳链为基础的简单基团的序列。组成生物体的主要元素包括:碳C、氢H、氧O、氮N、磷P、硫S、钙Ca等,这7种元素约占生物体质量的99.35%,其中C、H、O、N等4种元素约占96%。我们也可以对生命物质作更高度的抽象,即为C、H、O、N等4种基本元素的序列,其他元素的介入仅仅是增加了序列的复杂性。这种生命信息和生命物质从复杂到简单的跨越,可能就是物理学家们孜孜不倦地在追求的宇宙最根本的法则;同时,也体现了物质世界和精神世界的统一。

生命信息表达过程中的动力学特征给我们以下启示:

(1)信息存在的基础是序列的差异,系统中差异的有限性和差异的无限趋向性是生命系统稳定性和发展潜能的基础,是生物种系生存的基础,也是种系进化的基础。

(2)信息是差异复杂性的“突现”。由此,关于生命信息动力学可以抽象出一个重要的特征是:生命

信息表达的不守恒性和“涌现”。人类(3—4)万个基因可以涌现出无限的生命表象。

(3)在生命信息的表达系统中,熵增和熵减同时存在,而发挥主导作用的是熵减,即熵的逆向趋向性,使系统呈现不可还原的离散状态和非线性的不连续的过程。对非线性过程的研究意味着对传统的机械世界观的挑战,会给生命科学和医学带来深远的影响。

(4)混沌和秩序的边缘状态是生命过程动态演变的基础。研究表明,下述机制,包括:系统对初始条件的敏感性依赖,系统对反馈条件的敏感性依赖,阈限和临界群现象等,可能是生命活动和疾病过程的基本原理。

(5)建模和反馈机制的综合可能是研究和处理复杂系统的较好的技术路线。而简化过程仍然是最基础的原则,即在复杂中实现简化。

#### 参 考 文 献

- [1] Green D, Newth D. Towards a theory of everything? - Grand challenge in complexity and informatics. *Complexity International*, 2001, 8: 1—12.
- [2] Hanahan D, Weinberg R. The hallmarks of Cancer. *Cell*, 2000, 100: 57—70.
- [3] 沈岩,许琪. 人类疾病基因识别的研究进展. 中国科学院《2005'高技术发展报告》.北京:科学出版社,2005,84—89.
- [4] International Human Genome Sequencing Consortium: Initial sequencing and analysis of the human genome. *Nature*, 2001, 409:860—921.
- [5] Scott M. The natural history of genes. *Cell*, 2000, 100, 27—40.
- [6] Weng G, Bhalla U, Iyengr R. Complexity in biological signaling systems. *Science*, 1999, 284: 92—96.
- [7] Koch C, Laurent G. Complexity and the nervous system. *Science*, 1999, 284: 96—98.
- [8] Chen G, Goeddel D. TNF-R1 signaling: a beautiful pathway. *Science*, 2002, 296:1634—1635.
- [9] Zimmer G. Life after chaos. *Science*, 1999, 284: 83—86.
- [10] 李衍达,朱宗涵等译审, Baldi P, Brunak S 著. 生物信息学-机器学习方法.北京:中信出版社. 2003, 1—38.

### REVIEW OF CONCEPTS AND METHODOLOGY USED IN THE STUDIES ON COMPLEXITY ISSUES IN MEDICAL SCIENCES

Zhu Zonghan

(Beijing Municipal Health Bureau, Beijing 100053)

**Abstract** This paper reviewed the new concepts and methodology developed to manage and resolve the complexity issues in the scientific areas as well as in the medical areas. Bioinformatics has made the great contributions in studies on ge-

nomics and proteomics. The models and algorithms based on the non-linear dynamics have been shown the potential in understanding of biological evolution and ecology. It was also evidenced that those new scientific concepts and tools would be used to develop more comprehensive studies on the complexity issues of pathological courses and clinical practice, such as cancer, hypertension, and metabolic disorders and so on.

**Key words** medicine, complexity, bioinformatics, non-linear dynamics

·资料·信息·

## 关于受理国家杰出青年科学基金(外籍)申请项目的通告

为充分发挥海外科技人才资源优势,鼓励、吸引和稳定海外杰出科技人才在华工作,国家自然科学基金委员会在2005年启动国家杰出青年科学基金(外籍)资助工作,支持45岁以下具有较高学术水平和良好发展潜力的外籍华人青年学者全时全职在中国内地进行自然科学领域的基础研究。现将有关申请事项通告如下:

### 1. 申请者应具备的基本条件

(1)申请者是拥有其他国家国籍的华人学者,愿为中国科技事业发展和经济建设做贡献,并能够保证资助期内全时全职在中国内地开展基础研究(本通告的中国内地,系指我国除港澳台地区之外的省、自治区和直辖市)。

(2)申请当年1月1日未满45周岁。

(3)申请者已经为中国内地大学或研究单位正式聘用(惟一受聘单位,且在国外没有固定受聘职位),聘期覆盖本基金项目的执行期。

(4)申请者具有博士学位,现已被依托单位正式聘为教授级或相应专业技术职务。

(5)申请者具有较高的学术水平。曾在中国境外大学或研究机构从事科学研究,独立主持过若干重要的研究项目或课题,具有丰富的工作经验,研究成果已被国内外同行所认可,在本学科领域有一定的影响。

(6)申请者具有较强的组织能力与把握学术方向的能力。拟开展的研究工作方向明确,具有重要的科学意义和发展前景,属国际前沿且为国内所急需,可带动相关领域的发展或人才的培养。

(7)依托单位具有良好的支撑条件和发展环境,

能够为申请者提供必需的实验设备以及人力、物力,支付相应的工资报酬等。

### 2. 其他申请注意事项

(1)申请者须按照本通告和自然科学基金项目管理规定的要求,认真准备申请书及相关附件材料,通过依托单位申报。

(2)申请书使用通用《国家自然科学基金申请书》电子表格,按提示输入准确信息后打印输出。正文报告按照“国家杰出青年科学基金(外籍)申请书正文撰写提纲”的要求撰写。有关申请书和撰写提纲,请申请者自行到国家自然科学基金委员会网站(<http://www.nsf.gov.cn>)下载。

(3)本基金项目资助期限为4年,每人资助经费100万元(数学和管理科学领域资助70万元/人),请按《国家杰出青年科学基金项目资助经费管理办法》编制预算。

(4)依托单位学术委员会或专家组对申请人严格按照规定条件择优遴选并签署推荐意见。依托单位应对申请者的聘用情况及全时全职在华工作进行确认和负责。

(5)2005年申请截止日期为5月16日。依托单位将审核盖章后的纸质《申请书》和附件材料一式六份(至少一份为申请者签字原件)报送至国家自然科学基金委员会计划局,电子申请书通过国家自然科学基金网上申报系统提交。

本通告未尽事项,请参照《国家杰出青年科学基金实施管理办法》和《国家杰出青年科学基金项目资助经费管理办法》实施。有关问题请及时与国家自然科学基金委员会有关部门联系。