

文章编号: 1004-0374(2010)10-1035-08

计算系统生物学: 理论、方法及在药物研发中的应用

刁妍妍¹, 蔡超前², 蒋华良^{3*}, 李洪林^{1, 2*}

(1 华东理工大学药学院, 上海 200237; 2 华东理工大学信息科学与工程学院, 上海 200237;

3 中国科学院上海药物研究所药物发现与设计中心, 上海 201203)

摘要: 计算系统生物学是一个多学科交叉的新兴领域, 旨在通过整合海量数据建立其生物系统相互作用的复杂网络。数据的整合和模型的建立需要发展合适的数学方法和软件工具, 这也是计算系统生物学的主要任务。生物系统模型有助于从整体上理解生物体的内在功能和特性。同时, 生物网络模型在药物研发中的应用也越来越受到制药企业以及新药研发机构的重视, 如用于特异性药物作用靶点的预测和药物毒性评估等。该文简要介绍计算系统生物学的常见网络和计算模型, 以及建立模型所用的研究方法, 并阐述其在建模和分析中的作用及面临的问题和挑战。

关键字: 计算系统生物学; 生物模型; 建模方法

中图分类号: Q-1; Q-31 **文献标识码:** A

Computational systems biology: theory, method and applications in drug discovery

DIAO Yan-yan¹, CAI Chao-qian², JIANG Hua-liang^{3*}, LI Hong-lin^{1,2*}

(1 School of Pharmacy, East China University of Science and Technology, Shanghai 200237, China; 2 School of Information Science and Engineering, East China University of Science and Technology, Shanghai 200237, China; 3 Drug Discovery and Design Center, Shanghai Institute of Materia Medica, Chinese Academy of Sciences, Shanghai 201203, China)

Abstract: Computational systems biology is a multi-disciplinary field which aims at simulating the interactions of complex biological systems by integrating large quantities of biological data. In this process, it is essential to develop appropriate mathematical methods and simulation tools, which is also a main task of computational systems biology. The modeling of biological networks is of great value for further understanding the intrinsic properties of biological systems. Additionally, biological models have drawn increasing attention from pharmaceutical companies and research institutions for their application in target identification, toxicity assessment and so on. This review provides a brief summary of the available biological networks and modeling methods, elucidates their applications, and points out further challenges remaining in computational systems biology.

Key words: computational systems biology; biological models; modeling methods

近代生物学的研究倾向于个体的结构和功能, 人们做了大量实验以证实单个基因或蛋白质等生物分子如何在生命活动中发挥作用。在人类基因组计划的完成产生了海量数据的同时, 高通量技术及其在基因组和蛋白组等各种“组学”中的应用使生物信息的积累呈指数级增长。如何从纷繁复杂的单个生物数据中提取出有效信息, 进而从整体上理解生物体如何行使生命功能, 是生命科学发展的一个新

方向^[1]。为此, 科学家将计算机、数学及工程技术等研究方法引入到生物学域, 通过实验数据和计

收稿日期: 2010-05-04; 修回日期: 2010-05-19

基金项目: 国家重点基础研究发展计划(“973”项目)(2009CB918501); 国家自然科学基金项目(20803022)

*通讯作者 蒋华良, E-mail: hljiang@mail.shcnc.ac.cn, Tel: 021-50805873 李洪林, E-mail: hlli@ecust.edu.cn, Tel: 021-64250213

算模拟相结合建立起生物系统各部分相互作用的数学模型,以期定量地揭示生物体的整体功能。这一全新的策略使生物学研究跳出传统的方式,开启了从分子生物学到系统生物学(systems biology)的新时代^[2]。

建模方法和软件工具在系统生物学中的成功运用,使得计算系统生物学(computational systems biology)这一新兴学科受到越来越多的关注。计算系统生物学与计算生物学及生物信息学等紧密结合^[3],且更加关注建模方法、分析工具和软件系统等在系统生物学中的应用,其目的在于开发和运用更有效的数据整合方法和软件工具,使得生物学从实验水平上的描述性学科转为理论水平上的可预测性学科。

生物系统的复杂性在于各部分之间的因果互作和整合,为深入了解系统的作用和功能,既要明确个体之间存在什么样的关系,又要理解量的变化对这种联系产生怎样的影响,即在生物学研究中,应将定性的描述和定量的方法结合起来,计算模型在解决此问题的过程中尤为重要^[4]。计算系统生物学的最终任务是建立起各生物层次的数学模型,包括基因调节网络、蛋白质相互作用网络、代谢网络、信号转导网络及细胞循环模型,甚至各组织和器官的整体模型。这些生物系统的计算模型有助于人们从整体水平上理解生物体的构成和作用机制。此外,借助计算模型可以阐述疾病发生的机理和药物作用机制,也将为新世纪的药物研发提供了新的策略和方向。

1 生物网络和模型

在实际应用中,计算系统生物学主要包括以下几个方面(图 1)^[5]:首先,数据收集与分析,即将实验得到的或者文献中已有的生物数据的整合,对某一选定的生物系统进行了解和确定,描绘出该系统的结构,以此构造出系统模型的大体框架;其次,通过一定的计算方法对数据进行整合,建立起可预测的生物系统相互作用的网络模型,并对各组分进行定量描述;最后,将模型得到的结论与实验数据进行比较,对模型进行评价和修订,并根据修正后的模型作出预测或假设,设定和实施新的实验,重复上述步骤获得系统性的生物网络规律,为建立新的生物学理论框架奠定基础。

生物系统各组分之间存在多层次的非线性相互作用,在生物建模的过程中,人们将系统抽象为复杂的网络^[6],对生物系统进行直观的描述。以网络

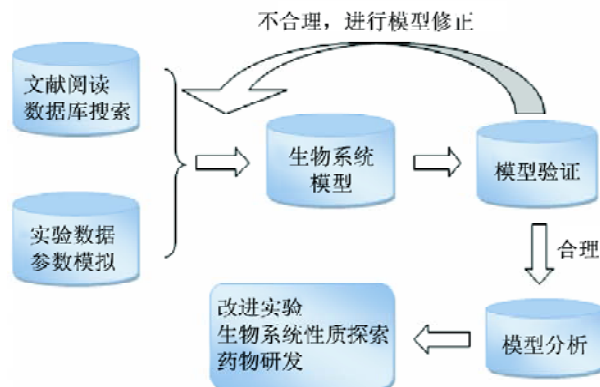


图1 生物系统建模的过程示意图^[5]

所需数据可从文献或实验中得到,缺失的参数用数学方法拟合;当模拟结果和实验数据一致时,模型被认为是可靠的;生成的模型可以进一步预测生物系统的性质

模型来解释生物现象,不仅可以描述生物系统的内在结构,也可以与数学方法结合,来预测系统响应外界变化的动态行为。

1.1 基因调控网络

基因调控网络(genetic regulatory networks, GRNs)在生命活动中起着重要作用,基因的状态决定了细胞分化、代谢过程和疾病的发生机理等。基因调控网络是DNA、转录因子、RNA、蛋白质及相关小分子相互作用的系统,并与代谢网络和信号转导网络等相关。对GRNs最基本的表述是基于逻辑结构的离散模型^[7],有向图是GRNs的最直接表述方法,节点表示基因和转录因子以及其他相关元素,边表示相互作用关系,通过边的不同定义描述基因调控的促进或抑制状态。Boolean网络将单个基因的活跃或停止状态用1和0表示,某时刻所有基因的 $0 \sim 1$ 矢量和构成系统的整体状态。为了揭示GRNs隐含的动态特征,也可采用微分方程等构建连续模型的方法;而聚类算法常用来对GRNs的结构特征和功能进行分析。

对GRNs的研究已逐渐应用到疾病机理研究、药物发现及基因调控的逆向重构等领域。GRNs相关的数据库主要是以功能注释的图形表示的^[8],如GeneNet、EcoCyc、RegulonDB及GeNet等。如何结合现有的生物学数据库,进一步挖掘GRNs所包含的信息,为复杂疾病研究提供指导,是系统生物学研究的重点之一。

1.2 蛋白质相互作用网络

通过蛋白质相互作用(protein-protein interactions, PPIs)的网络分析揭示细胞活动的功能是生命科学的

重要研究内容。在所有生物系统中, 蛋白质相互作用网络最为复杂, 且噪音最大^[9], 即使一个小的PPI网络, 包含的信息量也是巨大的。PPI是无向网络, 节点和边分别表示蛋白质及其作用关系。

由于高通量鉴别技术和计算机技术在PPI分析中的应用, 许多物种的蛋白质间相互作用得以有效识别, 并在此基础上构建了一系列数据库。与PPI相关的数据库现已有200多个, 如DIP、MINT、BioGRID、MPact及HPRD等。这些数据库为PPI网络的构建和分析提供了基础。鉴于快速增长的PPI数据库, 对PPI网络的分析成为系统生物学发展的推动力^[10]。基于PPI网络的规模, 分析方法大体分为两类^[11]: 对于蛋白质相互作用数量大的网络, 主要是分析具有重要生物学功能的网络拓扑结构性质; 而对于小规模网络, 主要是研究某一特定蛋白质与其他蛋白质的关系, 以拓展对其功能的认识。

1.3 代谢调控网络

代谢调控网络是由生物体内的代谢物、辅因子或酶参与的生化反应网络。近年来已积累了丰富的涉及到多个物种的代谢反应知识, 而研究的重点在于代谢网络的重构和相关结构和功能的分析。基于基因组的代谢网络重构是指根据基因组的功能注释信息, 结合相关的酶反应数据库、文献和实验数据, 对已有的代谢网络进行整合和修正, 建立起完整精确的针对特定物种的代谢反应网络^[12]。利用WIT、KEGG、EMP和MetaCyc等代谢通路数据库可方便地检索到参与某一网络的酶和代谢反应。基于不同的需求, 代谢模型用不同类型的有向或无向图表示, 进而进行后续的优化和分析。

代谢网络的构建常用基于约束的半定量方法, 如只需要部分反应物的浓度及酶的催化能力数据。流平衡分析(flux balance analysis, FBA)是常用的代谢网络建模和分析方法, 其本质是线性规划问题, 总体目的是寻求代谢网络的结构或过程最优化^[13]。对于基于基因组重构的代谢网络, 常有较大的规模。平均每个原核生物的重构代谢网络包含600个代谢物、650个基因和800个代谢反应, 而真核生物的代谢网络规模则更大^[14]。对于如此大规模的网络, 通常结合图论的方法分析其结构特征, 找出关键性的节点作为代谢控制点或潜在的药物作用靶点, 并从总体上探索生物体的功能。

1.4 信号转导网络

细胞信号转导是指细胞表面受体接受外界信号

刺激后, 通过一系列级联反应将胞外信号转化为胞内信号, 最终诱导基因表达等细胞应答反应的过程, 其在调节细胞内物质和能量代谢、控制细胞凋亡过程中起着关键作用。信号转导是多环节、多通路的高度复杂的可控过程, 并由此构成一个非线性网络。系统生物学期望建立各个细胞信号转导过程的模型, 找出起关键作用的蛋白, 并阐明其在疾病发生及基因调控中的作用。

近年来, 对信号转导网络的研究成为系统生物学的一大热点。关于信号转导网络的数据库也陆续建立起来, 如Signaling、及PID、BioCatra、SPAD、CSNDB和KEGG等综合性数据库也已扩增了信号转导网络数据。如何结合实验数据和计算机模拟方法, 建立通路中各组分随时间变化的定量模型是这一领域的普遍研究方法^[15]。人们试图将各种数学方法引入到信号转导网络的建模过程中, 以建立起能反映生物功能的精确的实时模型^[16], 并进行相关组分的功能分析。如可从BioModel数据库中搜索到已建立起来的信号转导模型, 相关数据也可从中获得。

2 建模方法

计算系统生物学是一个与计算模型和实验数据密切相关的学科, 数据模拟和计算机仿真有助于理解生物系统内在的演化过程和系统的动态行为, 并对系统进行简单预测, 以研究环境和系统之间的交互作用。数学模型试图以足够精确和细致的方法描述系统元素的状态、特征, 以及元素间的相互作用, 并在计算机上仿真系统的演化过程。

一种理想的形式化建模方法, 应该满足以下标准: 整合性, 即建模方法应尽可能描述系统内的所有进程; 直观性, 系统生物学是一个多学科交叉领域, 为了易于学科间的交流, 计算模型应以直观的形式描述系统的行为; 可扩展性, 随着实验和计算水平的进步, 人们对于生物系统行为的认识不断加深, 生物模型也需要不断的修饰或与其他模型整合, 建模方法须满足这一要求。常用的建模方法如下。

2.1 微分方程

尽管生物系统具有高度复杂性, 它们仍可被看做通过化学作用而相互联系的网络, 进而采用微分方程的数学方法进行分析。常微分方程^[17](ordinary differential equations, ODE)是描述动力学系统的常用方法, 由于ODE具有很好的数学理论基础, 因而

在系统生物学中得到广泛应用^[18-20]。如图2所示,系统中两个成分之间的相互作用(如激酶催化的底物磷酸化反应)可以简化为化学反应方程。从这一方程中,通过确定反应中的速率常数以及反应物的初始浓度,可以得到各组分随时间变化的微分方程,进而获得各组分与时间的关系曲线。用ODE模拟生化网络需要确定许多参数,不能由实验获得的参数则可通过数学模拟得到,这也是ODE的一个优势。相对复杂的生物网络将会涉及到多个ODE方程,这些方程组的求解可通过LSODA、CVODE等程序进行。SOSlib是一个基于ODE算法的开源计算机程序库,可用于分析和计算生化系统中的涉及到的ODE方程组。

ODE是定量研究细胞信号转导通路的基本方法,代谢网络中常用代谢控制分析方法(metabolism control analysis, MCA)也是建立每一代谢物的ODE方程^[21]。但ODE只能反映系统要素与时间的关系,并未考虑到物质转运和扩散等空间限制效应对反应的影响,对于同时存在时间和空间效应的系统则须建立偏微分方程(partial differential equations, PDE)。

系统组分的时空动力学性质可以通过如下PDE方程表示^[15]:

$$\frac{\partial C}{\partial t} = D \frac{\partial^2 C}{\partial x^2} - v \frac{\partial C}{\partial x} + R$$

(1)式中, C 表示各组分的浓度, t 和 x 分别是时间和空间变量, D 和 v 分别表示扩散系数和对流速率,而 R 代表反应物的生成或消耗速率。同ODE相比,尽管PDE更接近真实的生物系统性质,但PDE方法需要确定更多的参数,因而限制了其广泛应用。

需要指出的是,ODE/PDE方程作为一种确定性的建模方法,模拟的是系统的连续性变化行为。考虑到系统组分内在的随机性时,科学家人为地将随机因素引入到微分方程中,即随机微分方程^[22](stochastic differential equations, SDE)的方法。一般支持ODE建模的软件也同时支持SDE模型,如E-Cell和CellDesigner^[23]等。

2.2 随机进程代数

进程代数是用来描述并发进程的数学方法,各进程之间通过一定的规则共同协作,完成预期任

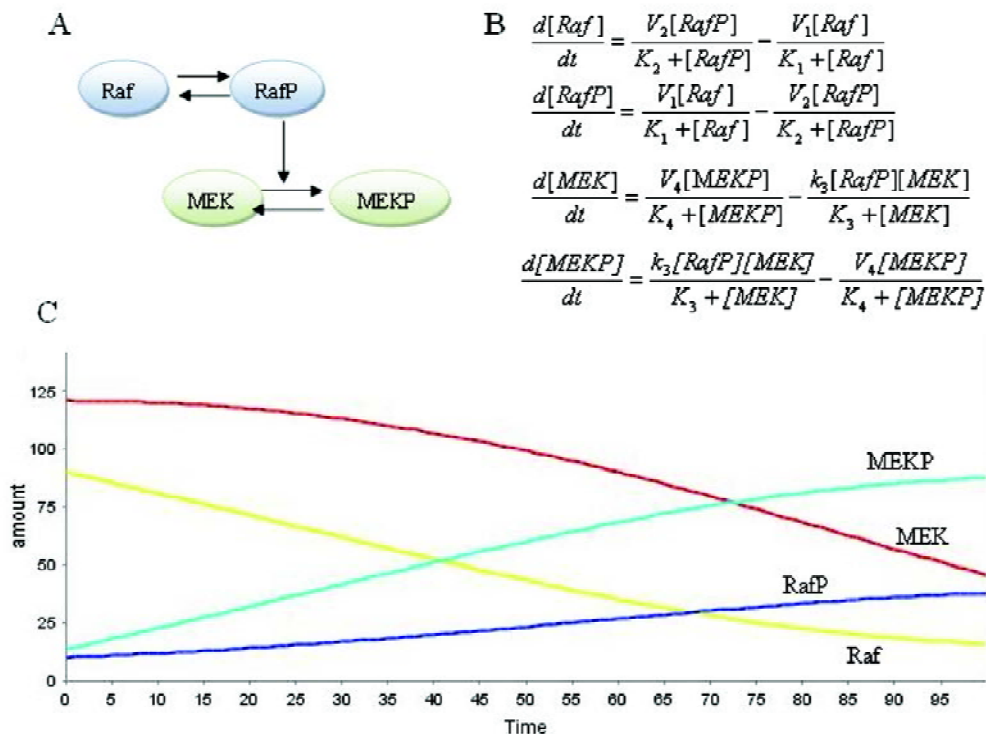


图2 ODE表示的简单生化反应

A: 磷酸化后的Raf能催化MEK的磷酸化,进而引起下游蛋白的一系列反应; B: 对反应中涉及到的各组分,建立其浓度随时间变化的ODE方程,同时需要确定各组分的初始浓度和涉及到的参数; C: 用CellDesigner对反应模拟的结果,即得到各组分浓度随时间变化的关系曲线,将各组分浓度变化情况直观地体现出来

务。对于生物系统中数量少的组分, 反应速率取决于分子间的碰撞几率^[24], 其动态性质往往具有随机性, 只有采用离散事件的描述才符合真实情况。生物系统中各部分反应同时发生, 且由于其相互作用, 各组分的状态处于不断变化中。所以, 用随机进程代数方法模拟生物系统, 更能反映系统内部的动态特征, 建立的模型更可靠。

2.2.1 随机 π -演算

π -演算是研究分布式并发系统的进程代数算法, 系统中的进程在互补的成对通道上同时作用, 通过传递信道来改变系统的状态。随机 π -演算通过加入随机因素定量地描述生物系统的随机性, 它依据概率分布对每个通道赋予一个速率, 使通道的通信能力不再是均等的, 这一方法已被应用到许多生物系统的建模过程中, 如Regev和Shapiro^[25]建模的信号转导过程是该方法的首次尝试。随机 π -演算不仅是从模拟整体系统状态的角度出发建立反应方程, 它还可以对单个反应组分的状态进行描述, 将这些组分通过一定规则连接起来, 即构成相互作用的网络系统。此方法将生物分子和特定的域看做计算进程(process), 而分子和域中的互补化学性质看做通信信道(channel), 化学反应和系统变化被看做通信过程和信道传递。

π -演算具有抽象的语法结构和语义, 理解起来有一定的难度。为了使 π -演算易于理解, 进而在系统生物学领域被广泛接受, Philips及其合作者发展了随机 π -演算的图形化表述方法^[26], 并在此基础上开发了SPiM和BioSPI等基于随机 π -演算建模框架的计算机仿真工具, 这些模拟工具有助于随机 π -演算的推广, 也为生物系统建模提供了多样化的方法。

2.2.2 性能评价进程代数

性能评价进程代数(performance evaluation process algebra, PEPA), 是系统生物学中用到的另一种进程代数方法^[27]。PEPA方法将模型描述为组分(components, 即生物分子)通过单个或多个活动(activities, 即生化反应)产生的交互作用。为了反应系统的多样性和反应的不确定性, PEPA将每个活动的持续时间看做服从负指数分布的随机变量。同随机 π -演算等其他进程代数方法相比, PEPA是对生物系统的更抽象的描述, π -演算以生物反应为基本单位来描述生物系统, 而PEPA侧重对反应组分状态的刻画。为了克服PEPA不能描述反应中物的计量关系等问题, Ciocchetta和Hillston^[28]发展了Bio-PEPA方法和相关工具。Bio-PEPA可以整合多个

分析技术, 对离散的随机系统或连续的确定的模型均具有较好的模拟能力。Priami等^[29]用Bio-PEPA方法建立了Gp130/JAK/STAT信号转导网络模型, 该模型得到的结果能很好地反应系统的行为, 也验证了Bio-PEPA在生物系统中的适用性。

2.3 Petri Net

Petri Net是用于信息处理系统建模的数学工具, 它可以将直观的图形化表述和完善成熟的数学分析方法很好地结合在一起, 因而在计算系统生物学领域有广泛的应用^[30]。Petri Net的不确定性、并行性、异步性, 以及对分布式系统的描述和分析能力使其在描述生物系统特性时有很大的优势^[31]。经典的Petri网包含两种节点, 即分别用圆形和矩形表示的位置(place)和变迁(transition), 节点通过有向弧(directed arcs)连接起来。在描述生物系统时, 位置表示涉及到的生物分子, 位置中的标记(token, 用点表示)代表分子数量或浓度; 变迁表示生化反应, 弧权表示化学计量系数(图3)。某时刻Petri Net的标识(marking)反应了生物系统的状态, 若有反应发生, 相应的变迁将被实施(fire), 网的标识会发生变化, 系统的状态也随之改变。

标准的Petri Net常用来定性分析生物网络的结构性质, 研究者将Petri Net进行了不同的扩展, 已解决更加复杂的系统生物学实际问题。时间Petri网(time petri net, TPN)基于网络的T-不变量性质, 可以判断变迁的可实现性^[32]; 随机Petri网(stochastic petri net, SPN)结合随机模拟算法, 通过反应时间延迟控制变迁的实施, 以定量刻画系统的随机特性^[33]; 混合Petri网(hybrid petri net, HPN)整合了微分方程模型, 能很好地模拟同时具有随机性和连续性的复杂生物系统。颜色Petri网(colored petri net, CPN)在

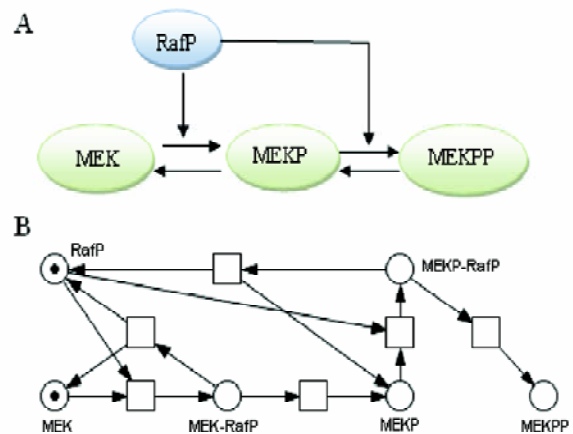


图3 Petri网表示的MEK磷酸化反应

HPN 的基础上将反应速率定义在变迁中, 增强了描述系统动态特性的能力^[34]。相应的 Petri Net 建模工具有 Snoopy、CPN Tools、Cell Illustrator 和 GON 等。

Petri Net 因其良好的直观性和数学分析兼容性, 在基因调控、代谢和信号转导网络的建模过程中都有很好的应用, 尤其在代谢疾病通路模型中的应用是一个成功的例子。如 Chen 和 Hofestädt 等^[35]用 Petri nets 建立了鸟氨酸循环的代谢模型, 利用此模型, 他们提出了治疗这种代谢紊乱疾病的方法, 如控制食物中蛋白摄取量和精氨酸饮食补充给药法等。

3 计算模型标准化形式

3.1 SBML

为了解决不同的计算系统生物学实际问题, 研究者用不同的方法构建了许多生物系统的计算模型。这些模型在格式定义和存储结构上存在很大的差异性, 同时, 计算工具的多样性也存在数据兼容性的问题。为了更好地交换和共享信息, 人们希望使用一种统一的语言描述生物系统的计算模型。系统生物学标记语言 (systems biology markup language, SBML)^[36]正是基于这种目的开发的计算机可读的、描述生物网络的文本语言形式。自 2000 年提出以来, SBML 得到了普遍的认可, 许多国际软件开发组织采用了该标准, 对其软件工具进行扩展以支持 SBML 格式^[37]。

SBML 能描述许多复杂的生物系统模型, 模型中的每个组分信息可以通过特定的具有层次的数据结构来表达。通过这种基于 XML 的描述, 其他的建模工具可以读取相应的 SBML 信息, 并通过特定的映射关系将模型转为自己的内部格式。因此, SBML 具有很好的兼容性和可操作性。LibSBML 是支持读写和验证 SBML 文件等操作的开源程序包, 为 SBML 的开发和应用带来很大的便利。

3.2 SBGN

SBGN (systems biology graphical notation) 是系统生物学中用来描述生命活动过程的图形标准^[38], 由欧洲生物信息学研究所 (EMBL-EBI) 和遍布世界各地的 30 个实验室于 2009 年联合发布。为了降低单一图形表述的复杂性, SBGN 定义了三种不同类型的图释: 过程图 (process diagram)、实体关系图 (entity relationship diagram)、活动图 (activity flow diagram), 以相互补充。这三种图释对生物系统刻画的详尽程度依次减弱, 各有侧重的图形表述适合不同的生物系统, 增强了系统生物学的表达能力。SBGN 中的

每个图形和符号都有明确的语义, 以确保对生物系统表述的准确性和高效性。目前发布的 SBGN 语言为 Level 1, 随着实际应用的不断深入和扩展, SBGN 也在不断完善中。

4 计算模型的应用和挑战

4.1 获取生物信息

人们用各有特点的方法建立各个水平的生物网络和计算模型, 是希望从这些模型中挖掘出有用的信息。对于计算模型的分析, 主要方法有两类: 基于网络拓扑结构的分析和基于系统组动力学特征的定量分析。对于网络拓扑结构的分析主要结合图论的方法以刻画生物系统结构的静态几何性质^[39], 对实验数据的要求较低, 如通过分析某节点的度分布 (degree distribution) 和介数 (between centrality), 来判断该节点在信息传输中的重要性, 又如通过聚类系数 (clustering coefficient) 反映系统组分的聚集状态。对于系统的定量分析主要是考查组分浓度或相关参数变化对系统的影响, 需要较多的精确实验数据支持。

通过模型分析得到的预测和假设可以帮助人们更好的理解生物有机体的结构和功能。在计算系统生物学的研究过程中, 常存在实验数据与模型预测结果的偏离。这种现象并不总是消极的, 从另一个角度讲, 它传达了这样一种信息: 对生物系统结构的现有理解是不完整的^[40], 计算模型中存在尚未被发现的缺失组分。从这一假设出发, 结合相应的实验验证, 有可能获得对生物体功能完整性的认识。同时, 利用已被实验证实的模型, 可以考查生物系统涌现出的一些新性质, 如鲁棒性 (robustness), 即系统对外界干扰所表现出的稳定性。

概括起来, 我们期望生物系统的计算模型具有以下作用: (1) 将复杂的生物系统网络直观地表示出来, 以助于理解各组分之间的相互作用; (2) 在计算机上模拟实验过程, 以节省时间、人力和物力; (3) 可以产生由于实验条件限制而无法获得的数据; (4) 预测生命现象并对实验数据进行合理解释; (5) 理性研究系统或某一过程如何行使其功能; (6) 鉴定系统过程的完整性, 找出缺失的成分。可以预见, 具有以上功能的系统模型有助于生物体结构和功能的认识, 将在生命科学的探索中发挥重要作用。

4.2 药物研发

传统的药物研发常常以单一靶标为基础, 而许多重大疾病是多靶标、多步骤作用的结果, 因而基于单靶标的药物研发不能全面反映药物与疾病的相

关性。系统生物学从整体上研究疾病通路的思想为药物研发提供了新的策略。生物模型在药物靶标发现和药物剂量确定中起着重要作用: 在疾病相关的调控网络中(与肿瘤发生相关的MAPK信号转导通路等), 关键性的控制点很有可能是药物治疗的潜在性靶标; 同样, 通过建立组织或器官的特异性代谢模型, 药物毒性测试会更加快速、高效; 更进一步, 如果建立多个器官组织的药物代谢作用模型, 可以更加快捷地确定用药剂量和周期。

计算系统生物学在药物研发中作用受到制药企业和医药研究机构越来越多的重视, 并产生了许多基于计算系统生物学的药物研发平台, 如整合基因干涉-药物-疾病信息的分析平台Connectivity Map^[41], 研究体外药物刺激对疾病细胞环境影响的模拟系统BioMap^[42], 具有准确功能注释的生物通路数据库分析平台IPA^[43], 以及进行“虚拟患者”(virtual patient)模拟的病理学预测平台PhysioLab Platform^[44]。这些药物研发平台已被许多国际制药企业成功应用到早期靶标验证、药物活性测定和后期临床药理学评价当中, 可更好地预测候选物的临床作用和潜在的不良反应^[5]。

4.3 面临的挑战

计算系统生物学是一个多学科交叉的新领域, 各思想的交汇、各方法的应用使计算系统生物学处于一个“百家争鸣”的繁荣时代。不同领域研究者的关注和努力促进了系统生物学的理论发展和实际应用, 但同时, 面临的挑战也是多方面的:

(1) 数据的分析及整合。系统生物学是基于大量生物学信息的整合学科, 数据的鉴定和整合是首要问题。从不同数据库中找出模型的重要组分, 建立合理的系统网络需要花费大量的时间; 不同实验方法得到的数据也有很大的区别, 在对参数确定过程中需要仔细鉴别; 另一方面, 将不同类别的数据信息, 如生物分子三维结构、环境条件和功能文本注释信息^[9], 整合到模型中更是一个尚未得到解决的难题。

(2) 建模方法的局限性。ODE及随机代数等建模方法并不是为了系统生物学而设计的, 在模拟生物系统时难免存在不足。尽管研究者对不同方法进行了扩展, 以更好地适应生物系统的特性, 但也仅是对生物系统模拟的不同程度的近似。寻找更加适合生物系统的建模方法将是计算系统生物学的长期任务。此外, 现有的建模方法多关注于生物组分量与时间的变化关系, 如何同时考虑系统的时间和空

间模型将是下一步的研究方向^[45]。

(3) 生物模型的交换和整合。现有的生物模型是采用不同的方法、基于各自特点的实际生物学问题建立的。不同格式、规模的计算模型在交流和整合上存在很大的障碍; 同时对于某一具体的模型, 随着人们认识水平的加深, 也需要进行结构上的修正和规模上的扩展。尽管SBML、SBGN标准试图解决模型表述之间的差异性, 如何实现已有模型和方法与SBML、SBGN之间的自动转化, 仍是一个需要解决的问题。

(4) 从数据到图形。ODE等数学方法仅仅是描述生物量的变化, 并不能直接以图形的方式展现生物系统的结构特性。若可实现生物定量数据与图形的结合与自动转化, 将能更好地刻画生物系统结构与动态性质之间的关系。

5 结语

计算系统生物学是后基因时代的新兴学科, 已成为生命科学领域的热点和前沿, 它建立在大量的生物数据基础上, 以计算模型研究为手段, 系统地分析生物整体的性质。总体说来, 这一学科发展时间较短, 还需要进一步发展新的研究方法和软件工具。在未来, 计算系统生物学的发展还面临着多方面的挑战, 同时, 其发展的空间也是十分广阔的。计算系统生物学的思想和研究方法冲击着生命科学的各个领域, 结合自身优势进行系统生物学的相关研究, 实现理论研究到实际应用, 才能积极应对这种趋势。

[参 考 文 献]

- [1] Hartwell LH, Hopfield JJ, Leibler S, et al. From molecular to modular cell biology. *Nature*, 1999, 402(6761): 47-52
- [2] Hood L. Systems biology: Integrating technology, biology, and computation. *Mech Ageing Dev*, 2003, 124(1): 9-16
- [3] Kitano H. Computational systems biology. *Nature*, 2002, 420(6912): 206-10
- [4] Deville Y, Gilbert D, Van Helden J, et al. An overview of data models for the analysis of biochemical pathways. *Brief Bioinform*, 2003, 4(3): 246-59
- [5] Rajasethupathy P, Vayttaden SJ, Bhalla US, et al. Systems modeling: a pathway to drug discovery. *Curr Opin Chem Biol*, 2005, 9(4): 400-6
- [6] Barabasi AL, Oltvai ZN. Network biology: understanding the cell's function organization. *Nat Rev Genet*, 2004, 5(2): 101-13
- [7] Karlebach G, Shamir R. Modelling and analysis of gene regulatory networks. *Nat Rev Mol Cell Biol*, 2008, 9(10): 770-80

- [8] de Jong H. Modeling and simulation of genetic regulatory systems: a literature review. *J Comput Biol*, 2002, 9(1): 67–103
- [9] Heath AP, Kaviraki LE. Computational challenges in systems biology. *Comput Sci Rev*, 2009, 3(1): 1–17
- [10] Yosef N, Ungar L, Zalekvar E, et al. Toward accurate reconstruction of functional protein networks. *Mol Syst Biol*, 2009, 5: 248
- [11] Wu J, Vallenius T, Ovaska K, et al. Integrated network analysis platform for protein-protein interactions. *Nat Methods*, 2009, 6(1): 75–7
- [12] Kim TY, Kim HU, Lee SY. Data integration and analysis of biological networks. *Curr Opin Biotechnol*, 2010, 21(1): 78–84
- [13] Murabito E, Simeonidis E, Smallbone K, et al. Capturing the essence of a metabolic network: a flux balance analysis approach. *J Theor Biol*, 2009, 260(3): 445–52
- [14] Oberhardt MA, Palsson BØ, Papin JA. Applications of genome-scale metabolic reconstructions. *Mol Syst Biol*, 2009, 5: 320
- [15] Goldbeter A. Computational approaches to cellular rhythms. *Nature*, 2002, 420(6912): 238–45
- [16] Phar T. *Computational Biology: Issues and Applications in Oncology* [M]. New York: Springer-Verlag, 2009: 227–42
- [17] Gilbert D, Heiner M. From petri nets to differential equations—an integrative approach for biochemical network analysis. *Lect Notes Comput Sc*, 2006, 4024: 181–200
- [18] Sible JC, Tyson JJ. Mathematical modeling as a tool for investigating cell cycle control networks. *Methods*, 2007, 41(2): 238–47
- [19] Kell DB. Metabolomics, modelling and machine learning in systems biology—towards an understanding of the languages of cells. *FEBS J*, 2006, 273(5): 873–94
- [20] Eungdamrong NJ, Iyengar R. Modeling cell signaling networks. *Biol Cell*, 2004, 96(5): 355–62
- [21] Golightly A, Wilkinson DJ. Bayesian inference for stochastic kinetic models using a diffusion approximation. *Biometrics*, 2005, 61(3): 781–8
- [22] Reinker S, Altman RM, Timmer J. Parameter estimation in stochastic biochemical reactions. *Syst Biol: Stevenage*, 2006, 153(4): 168–78
- [23] Funahashi A, Matsuoka Y, Jouraku A, et al. CellDesigner 3.5: A versatile modeling tool for biochemical networks. *P IEEE*, 2008, 96(8): 1254–65
- [24] Ciocchetta F, Hillston J. Process algebras in systems biology. *Lect Notes Comput Sc*, 2008, 5016: 265–312
- [25] Regev A, Shapiro E. The π -calculus as an abstraction for biomolecular systems [M] // Ciobanu G. *Modelling in molecular biology*. Berlin: Springer-Verlag, 2004: 219–68
- [26] Phillips A, Castagna G, Castagna G. A graphical representation for biological processes in the stochastic π -calculus. *Lect Notes Comput Sci*, 2006, 4230: 123–52
- [27] Geisweiller N, Hillston J, Stenico M. Relating continuous and discrete pcpa models of signalling pathways. *Theor Comput Sci*, 2008, 404(1–2): 97–111
- [28] Ciocchetta F, Hillston J. Bio-pcpa: An extension of the process algebra pcpa for biochemical networks. *ENTCS*, 2008, 194(3): 103–17
- [29] Priami C, Back RJ, Petre I. *Transactions on Computational Systems Biology: XI* [M]. Germany: Springer Berlin/Heidelberg, 2009: 90–115
- [30] Murata T. *Petri nets: Properties, analysis and applications*. P IEEE, 1989, 77(4): 541–80
- [31] Peleg M, Yeh I, Altman RB. Modelling biological processes using workflow and petri net models. *Bioinformatics*, 2002, 18(6): 825–37
- [32] Duguid A, Gilmore S, Guerriero M, et al. Time petri nets for modelling and analysis of biochemical networks. *Fund Inform*, 2005, 67: 149–62
- [33] Napione L, Manini D, Cordero F, et al. On the use of stochastic petri nets in the analysis of signal transduction pathways for angiogenesis process [C] // *Computational methods in systems biology: International Conference CMSB 2009*. Berlin: Springer-Verlag, 2009: 281–95
- [34] Lee DY, Zimmer R, Lee S, et al. Colored petri net modeling and simulation of signal transduction pathways. *Metab Eng*, 2006, 8(2): 112–22
- [35] Chen M, Hofestadt R. A medical bioinformatics approach for metabolic disorders: biomedical data prediction, modeling, and systematic analysis. *J Biomed Inform*, 2006, 39(2): 147–59
- [36] Hucka M, Finney A, Sauro H, et al. The systems biology markup language (sbml): a medium for representation and exchange of biochemical network models. *Bioinformatics*, 2003, 19(4): 524–31
- [37] Hucka M, Finney A, Bornstein B, et al. Evolving a lingua franca and associated software infrastructure for computational systems biology: The systems biology markup language (sbml) project. *Syst Biol*, 2004, 1(1): 41–53
- [38] Le Novère N, Hucka M, Mi H, et al. The systems biology graphical notation. *Nat Biotechnol*, 2009, 27(8): 735–41
- [39] Aittokallio T, Schwikowski B. Graph-based methods for analysing networks in cell biology. *Brief Bioinforma*, 2006, 7(3): 243–55
- [40] Ideker T, Galitski T, Hood L. A new approach to decoding life: Systems biology. *Annu Rev Genomics Hum Genet*, 2001, 2: 343–72
- [41] Materi W, Wishart DS. Computational systems biology in drug discovery and development: methods and applications. *Drug Discov Today*, 2007, 12(7–8): 295–303
- [42] Berg EL, Kunkel EJ, Hytopoulos E, et al. Characterization of compound mechanisms and secondary activities by biomap analysis. *J Pharmacol Toxicol Methods*, 2006, 53(1): 67–74
- [43] Ganter B, Abramovitz DL. Application of toxicity and biomarker workflows within ingenuity pathways analysis (IPA) results in efficient evaluation of compound toxicity and safety. *Toxicol Lett*, 2007, 172(1): 45
- [44] Kansal AR, Trimmer J. Application of predictive biosimulation within pharmaceutical clinical development: examples of significance for translational medicine and clinical trial design. *Syst Biol*, 2005, 152(4): 214–20
- [45] Zhu H, Huang S, Dhar P. The next step in systems biology: Simulating the temporospatial dynamics of molecular network. *Bioessays*, 2004, 26(1): 68–72