

一个新兴的交叉学科： 系统生物学

孙之荣 吴雪兵
清华大学

关键词：系统生物学 交叉学科

前言

什么是系统生物学

一直以来，生物体都被认为是最典型的复杂系统^[1]，人们对于这种系统的行为已经非常熟悉，但是对这些行为背后的工作机制却知之甚少。系统生物学致力于解决这一问题，它将帮助人们从系统水平上理解生物体的结构和功能。

系统生物学，它最初的定义，或者说狭义的定义是由诺伊·胡德（Leroy Hood）提出的，指在基因组、转录组、蛋白质组和代谢组等复杂性的全局分析的基础上，建立生物系统的模型，从中归纳假设，并利用这些假设对生物系统进行反复的、整合性的干扰响应研究^[2]。这一思想和研究方式取得了显著的成效，并被迅速推广，一些其他领域的学者的研究工作也不断拓展这一领域，但这些研究的基本思想是一致的。一个更加广义的定义可以概括它们的共同特征：系统生物学是研究一个生物系统中所有组成成分的构成，以及在特定条件下这些组分之间的相互作用关系的学科^[3]。系统生物学是在分子生物学基础上发展起来的，它以整合为特征，综合了系统分析、高通量平台、计算机模拟、数据整合与信息发掘等领域的知识，来获取生物系统的结构和功能属性，在系统层次

上理解生物。

系统生物学的交叉属性

系统生物学是一个典型的跨学科交叉领域，而又不仅是若干学科的简单借鉴。可以从以下三个方面来阐述系统生物学的交叉属性。

首先，在理论思想上，系统生物学以系统学为指导，以整合为特征，这一点与系统和控制论指导下的生物控制论一脉相承。二者均从系统水平理解、控制和设计生物系统为根本目的，总体上采用控制论的理论和方法建立模型，研究系统行为。但是二者的区别也是显而易见的，这一点将在下文分析。

其次，在技术手段上，系统生物学得益于分子生物学发展起来的高通量技术平台，包括大规模DNA¹测序技术、基因芯片技术和多参数高速细胞分类仪等。

最后，在研究方法上，系统生物学依托于生物信息学（计算生物学）发展起来的大型数据库、软件模拟和分析平台、数据整合与信息发掘技术、结构/功能预测算法等。

上述三个方面又是由三个更大的学科相互结合产生，即生物学、系统论和控制论、计算机科学与技术。这三个学科和领域相互作用，产生了上述三个交叉学科和技术领域，具体地说，控制论与生物学的结合产生了生物控制论，生物学与计算机的结合出现了生物信息学，计算机技术和控制论的融合应用则是现代

¹ deoxyribonucleic acid, 脱氧核糖核酸

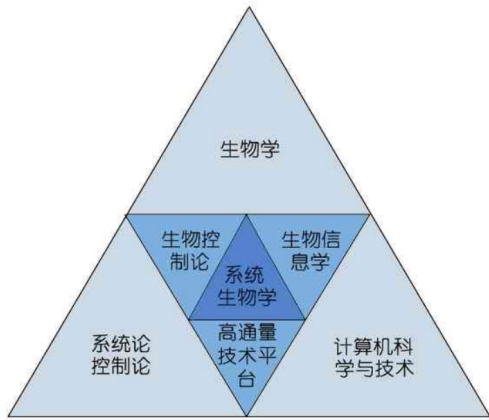


图1 系统生物学内学科间的相互关系

自动化的高通量测量技术仪器设计和制造的依托。

图1刻画了这种关系。图中相邻三角形之间有着借鉴、依赖、派生和发展的关系。比如“生物控制论”处于“生物学”和“系统论控制论”之间，较形象地表明了三者间的关联。其他三角形间均如此。

需要指出的是，图1并不表示系统生物学

由图中所列举的那些学科和技术领域组成，而是突出它们在某一方面的关联。实际上，系统生物学涉及面极其宽泛，在下文中我们将会看到，系统生物学的发展，需要生物学、数学、控制论、计算机科学、物理学和化学等各个学科的支持，需要生物学家、数学家、计算机科学家和工程师等的通力合作，系统生物学的发展也将极大地促进这些学科的建设。

系统生物学工作框架

有些学者认为，生物学本质上是信息科学^[4]，编码在DNA中的信息沿着如下信息流流动：

DNA → mRNA² → 蛋白质 → 蛋白质相互作用网络 → 细胞 → 器官 → 个体 → 群体

系统生物学的重要任务就是尽可能获取每一个层次的信息，并且对它们进行整合。系统生物学的目的就是要研究并揭示生物信息的规律。

从这个角度，可以将系统生物学的研究理解为对信息反复进行处理的过程。其框架如图2所

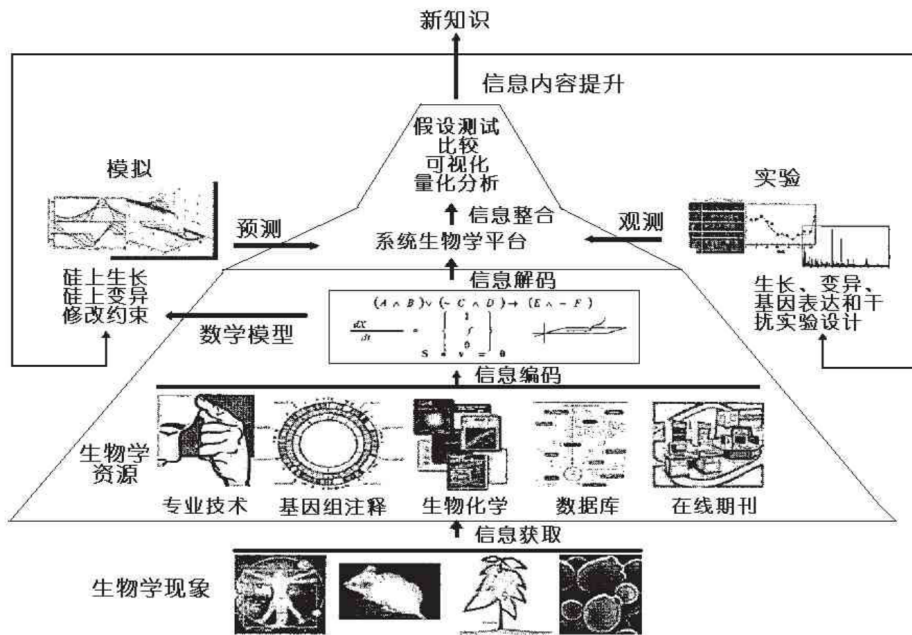


图2 系统生物学框架（修改自参考文献^[5]）

² messenger RN, 信使核糖核酸

示。处理过程包括：从生物现象中获取信息；对生物资源进行编码；用数学工具解码；对计算结果和实验结果进行整合；最后从量化和系统性的分析中得到新的知识。

下面将在不同方面进一步阐述这一框架。首先介绍生物控制论，它是系统生物学出现之前，人们将系统与控制引入生物学研究的尝试。生物控制论的研究对于系统生物学的发展具有许多启示和帮助。

从生物控制论到系统生物学

控制论诞生之初，就与生物学紧密相关。在维纳（Weiner N.）发表的《控制论——关于在动物和机器中的控制与通信（Cybernetics or Control and Communication in the Animal and the Machine）》一书中，这位伟大的控制论之父认为，工程学中所采用的系统与控制的理论和方法同样适用于生物学研究^[6]。此后，维纳本人和一大批学者在这一领域进行了大量研究，逐步建立了生物控制论这一学科。

生物控制论以控制论的理论和研究方法研究生物系统的调节控制和信息处理过程，其目的在于建立能够反映人体和动物功能的模型和理论。这种模型和理论中的逻辑和原理，与有机体本身中起作用的原理是相同的。生物控制论的研究对象集中于生理水平，主要为病理学和神经生理学，如血糖代谢、胰岛素的调节、体温调节等提供模型。这些生物控制论的实践为系统学和控制论在系统生物学中的应用提供了认识和经验，包括如下内容：

- （1）对生物系统鲁棒性的理解
- （2）对生物系统的定量研究
- （3）大规模生理系统的模拟
- （4）药物动力学、细胞动力学研究

系统生物学区别于以上这些系统性尝试的最大不同，首先在于系统生物学有可能将系统水平的理解直接建立在分子水平如基因、蛋白

质等之上，而过去的尝试不可能建立系统水平描述和分子水平知识之间充分的联系。因此，尽管这不是第一次从系统水平进行理解，却是第一次有机会从分子水平到系统水平建立坚实的知识结构进而理解生物学系统^[7]。其次，上述理论和模型的发展，都遇到某些局限，就是对于组成这些系统的单元的分析性工作太少，支撑这些理论和模型的数据不够充分，无法收集到足够的关于研究对象的信息。这主要由于受到当时分子生物学发展水平和试验测量技术的限制。这两个方面的突破，正是分子生物学的发展带来的。

从分子生物学到系统生物学

系统生物学现在能够从分子生物学层次来研究和发育，直接原因是由于单个基因、单个蛋白质的研究并不能解释基因的功能，必须在众多基因与蛋白质作用中了解其功能。现在大规模基因组计划的实现，加快了分子生物学新时展的发展。

1953年，DNA双螺旋结构的提出，标志着分子生物学的诞生。这一事件被认为是20世纪最重要的科学发现之一，其重要意义在于，首次将生物学研究建立在分子基础之上，使生物学研究有了坚实的理论基础，而这正是系统生物学与生物控制论的最本质区别。分子生物学家使用还原论的方法，将生物系统逐级分解、“还原”，并发展了各种分子生物学试验技术和测量仪器与设备。这些研究使我们能够深刻理解遗传、进化、发育和疾病等基本过程。但是这些研究集中在识别基因以及认识它们的表达产物的功能，而生物系统的功能蕴藏在系统的整体结构和各种组分的相互作用之中，因此即使知道了所有成分的结构和功能，也不足以解释复杂的生物系统。

于1990年启动的人类基因组计划（Human Genome Project, HGP）标志着分子生物学更大的

新发展,也是分子生物学发展的转折点。这是生物学第一次将整个生物系统的所有元素加以定义,并规范了在数据存储格式。从此,对生物学的研究开始由分解转向整合。在HGP的带动下,出现了一系列组学(Omics),如转录组学、蛋白质组学、代谢组学和相互作用组学等,这些专门领域的研究为系统生物学提供了系统元素列表,逐步把分子生物学推向系统生物学。

高通量平台是分子生物学研究的技术平台,为系统生物学发展提供了强大的工具。系统生物学虽然综合了数学、计算机和控制论等生物学之外的科学,其本身也是一门实验性科学,在它最初的定义中尤其强调重复的实验验证环节。目前的系统生物学刚刚从分子生物学发展而来,并没有十分成熟的理论基础,需要通过大量的实践,逐步建立各种模型及其理论描述,而其研究对象为各层次的生物系统,这一过程必须结合具体对象,进行具体实验,否则只能回归到对于一些现象的描述。系统生物学研究系统属性时,所要求的全局数据集的获取与分析,各种干扰实验的实施,都离不开高通量、高精度的实验仪器与设备。

在过去20年的时间里,分子生物学发展了一系列的基因组、蛋白质组、代谢组和细胞层面的测定和分析技术,以1986年第一台DNA自动测序仪为标志,陆续开发了其他仪器和技术,如PCR³技术、DNA合成仪、DNA芯片和cDNA⁴阵列、质谱分析、核磁共振技术、蛋白质芯片和用于蛋白质质量研究的毛细管分离(capillary separations)技术,以及其他用于测量蛋白质相互作用的仪器和技术。上述技术和仪器都向着小型化、综合化、并行化和自动化的方向发展,并逐渐结合先进的微流体和纳米技术^[8]。

上述技术仍将在系统生物学中广泛使用,同时系统生物学也对实验设备提出了新的挑战^[9]。

比如,为了建立生物系统的模型,需要确定大量的参数,而且从数学建模的角度考虑,要求同时测量。又比如,细胞内大量生化反应在时间尺度上相差甚远,但在功能上又不能忽略其中的某一个反应,因此需要使仪器在测量范围和精度上都满足要求。另外,在空间尺度上,广泛使用的培养皿等仪器易受扩散时间和容量的影响,为了获得与细胞内反应事件相对应的应答,仪器本身需要和细胞处在同一尺度,而且应当具有同样的灵敏性,才能保证精度。上述问题都极具挑战性。

对于系统生物学来说,除了测量各种需要的参数外,更重要的是如何对其进行控制。在信号转导、代谢通路等模型研究中,分子水平的传感器和执行机构将发挥越来越重要的作用。这些方面的发展,也需要各领域的通力合作。

系统科学与控制论在系统生物学中的应用

系统和控制理论自诞生以来,在制造、电子和通讯等领域都取得了巨大成就,甚至包括像经济和社会这样的系统中都进行了较为深入的研究。生物控制论则是其在生物领域的发展,如前所述,它为系统和控制理论在生物学中的应用积累了经验,获得了生物系统性的一些初步特征。但是,这些研究基本上集中在生理水平,而且随着基因组学和高通量技术的发展,出现了新情况和新问题,系统和控制理论有了更大的研究空间,可以进行更深入的研究。这些研究包括生物系统的系统辨识与建模、系统涌现属性、生物鲁棒性以及调控网络机制的研究等诸多方面。由于篇幅的关系,不可能逐一展开,下面将重点介绍系统辨识、参数敏感性、涌现

³ Polymerase Chain Reaction, 聚合酶链反应

⁴ 互补DNA

属性和鲁棒性这四个方面。

系统辨识与逆向工程

系统辨识利用系统的输入输出建立系统的数学模型。在系统生物学中,对系统施加各种不同的输入,然后测量相应的基因和蛋白质网络中某些可测量的变量——如蛋白质浓度和RNA转录量等,再利用控制论中系统辨识和实现理论建立系统模型,供进一步研究,而在实际研究中,往往可以直接利用已经公布的大量数据进行辨识。

系统辨识一般包括系统结构辨识和参数辨识。首先,需要确定系统的结构,即明确系统所有组成成分和这些组分之间的相互作用关系;其次,需要测定这些组分及相互作用关系的有关参数,如组分浓度和相互作用强度等,以建立精确模型。结构辨识的基础工作是发现和预测相互作用关系,包括基因相互作用和蛋白质相互作用,都可以通过实验测定。但目前,很多作法是先通过对已有的相互作用数据库使用一些预测算法来发现新的相互作用关系。这时,通路数据库和相互作用数据库(如KEGG^[10]等)就显得较为重要。除此之外,也有直接利用实验数据预测网络结构的,这一类方法使用DNA芯片及其他高通量测量技术获取大量数据,使用聚类分析方法推断网络结构,其优点是能够处理大规模数据。然而,这类方法一般需要结合启发性思路,并且需要大量的计算,而且预测效果有限,大多数情况只能按照某种条件对组分进行分类。但是,如果与第一种方法结合使用,并做某些扩展,将可以发挥更大的作用,如进行基因调控网络的推导^[11]。

只有参数确定的网络,才能进行评判和检验。系统参数辨识基于大量直接或间接的测量数据,很多参数是无法直接测量的,除了技术上的原因外,从系统的角度,生物体本身就

很多状态变量是不可观测的。参数值的拟合与优化,依赖于一些算法,如基因算法、模拟退火及神经网络等人工智能算法,此外还需要专门的软件。参数的估计也应该在实验及仿真过程中不断优化,使得模型与实验结果较好地吻合。在这一过程中,有三点需要特别注意:首先,能够产生相似行为的参数集乃至系统结构可能不惟一;其次,不同参数需要的精确程度不同,即敏感性不同;第三,参数优化算法应该尽可能发现局部最优,而不是全局最优^[7]。

参数敏感性

在辨识系统参数时,参数敏感性的估计十分必要。在复杂系统中,存在大量的参数,精确辨识所有参数工作量极大,而系统的行为往往只取决于少数的关键参量。很多参数并不需要精确确定,因为系统对这些参数的变化具有较强的鲁棒性,也就是不敏感。这类系统的功能一般蕴含在系统结构中^[12],而不是具体的组分与参数,比如细菌的趋化性。而另外一些参数则不同,这些参数的细小变化都有可能导致系统行为的极大改变。一般而言,这些参数是系统的关键所在。找到这些敏感参数,不仅对于有助于理解系统的本质,而且可以大量减少系统辨识所需要的数据量和工作量;同时,当系统发生故障(如各种疾病)时,一般是这些参数发生了变化,可以据此迅速地诊断和有针对性地治疗。

以往参数敏感性分析都只关注单个参数,即在其他参数不变的情况下,观察该参数改变时系统输出变化的剧烈程度。对于生物体而言,这种方法具有局限性,因为系统的性能并不取决于单个参数的叠加,而是多个参数的非线性作用的结果。因此,需要能够一种全局的多参数敏感性分析方法。Zi ZK等借鉴水文领域的多参数敏感性分析方法(MPSA⁵)分析了Jak-Stat⁶通路,确定了该信号通路中的几个关

⁵ Methods of Protein Structure Analysis

⁶ JAK: Janus kinase, 激酶; STAT: Signal Transducer and Activator of Transcription, 信号转导和转录激活子

键组分和作用^[13]。

涌现属性

涌现 (emergence) 是非线性系统的一个本质特征, 是指由大量元件通过相互作用而产生的单个元件所不具有的行为或特征。这些行为和特征不可能通过单个元件作用的叠加而得到, 表现为整体大于部分和。生物系统是一个复杂的非线性系统, 它的很多行为都表现出涌现的特征。这也是生物学研究从分析走向综合的原因之一。

在细胞信号转导系统中, 信号传导通过共同的中间物质产生关联形成信号网络。这些网络内物质存在各种各样的相互作用, 很容易产生一些较复杂的行为。比如时间尺度上跨越很大的多个信号在一个通路中的整合; 一些输出取决于输入的强度和持续时间长短; 自我维持的反馈环; 在两个状态间转换有精确的阈值等等。这些行为也不是彼此无关的, 整个网络因为这些涌现属性而可能具有宏观上的一些更复杂的行为。比如, 知识、信息和记忆可能就储存在这些信号网络中^[14]。这些行为都不能孤立地从网络中物质的性质去理解。即使是在非常简单的生物系统中, 这种涌现属性也是存在的^[15]。

鲁棒性

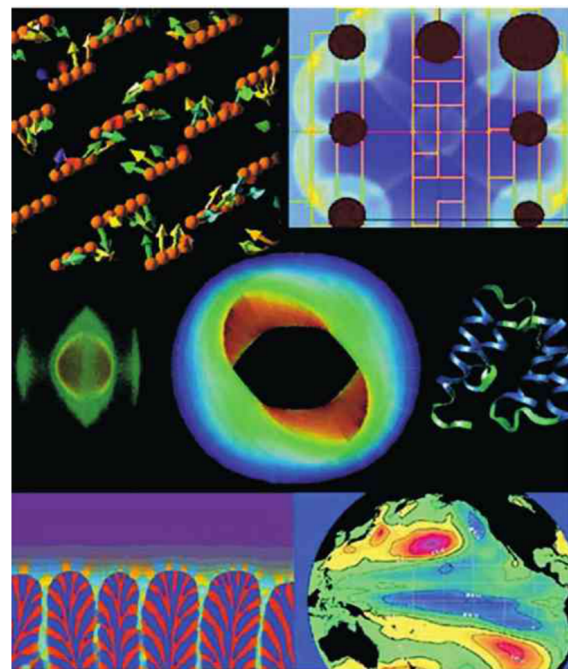
鲁棒性是生物系统主要的涌现特性之一, 也是目前系统生物学研究热点, 对于我们理解复杂疾病原理及其治疗设计极为重要。在控制论中, 鲁棒性是指系统在内外干扰下保持自身功能的能力, 它使得系统能够用不可靠的元件在不可预知的环境中稳健地运作。鲁棒性通常通过四种机制获得: 系统控制、冗余、模块化和解耦 (或结构稳定性)。鲁棒性在生物系统中普遍存在, 上述机制在生物体中都可以找到^[7]。

目前生物鲁棒性研究主要集中在生物系统鲁棒性现象、鲁棒性起源、鲁棒性的“折中” (tradeoffs)、疾病的鲁棒性观点和鲁棒性的理

论化研究。对于生物系统中鲁棒性的起源尚无定论, 北野武 (Kitano) 等认为鲁棒性的起源在于进化, 鲁棒性是进化着的复杂动态系统的内在属性——各种导致有机体鲁棒性的机制实际上都推动了进化, 进化也选择了鲁棒的性状^[16]。

尽管如此, 鲁棒性也是“双刃剑”, 鲁棒性能够增强系统对于常见干扰的适应性, 但对于新的未知干扰, 系统却极端脆弱。在鲁棒性与脆弱性 (fragility)、性能与资源需求之间, 存在“折衷” (tradeoffs), 系统鲁棒程度的增加, 必然同时引入脆弱性, 增加对资源的需求。比如, 相对于怀特兄弟的第一架飞机, 现代大型商用客机通过引入复杂的控制系统实现了对大气变化很强的鲁棒性, 这是以极大的资源需求为代价的, 同时也引入了不少潜在的致命弱点, 比如使社会电力供应紧张。又如, 细菌趋化性中, 负反馈能够提高细菌跟随化学梯度的能力, 使其对外界化学浓度的改变具有鲁棒性, 但是如果如果没有负反馈的话, 细菌会游动得更快, 鲁棒性的代价是游动速度的降低。

很多复杂疾病都可以从鲁棒性伴随脆弱性



的角度来理解。比如糖尿病可以认为是系统对能量供应相对不足、接近饥饿的状态获得了鲁棒性，但异常地被过度营养而低能量需求的生活方式干扰^[17]。此外生物体正常的鲁棒性可能会被疾病利用，从而使机体自身调节和药物治疗失去效果。如抗药性是由MDR1⁷和其他基因的正向调节（upregulation）产生的，这些基因的产物将有毒化学物质排出细胞，在正常情况下保护生物体的安全，但是在癌症中被肿瘤用来保护恶性细胞，使其具有抵抗药物的能力。对于这些疾病的治疗，也应该从鲁棒性的角度来设计，即寻找伴随这些鲁棒性的弱点，重新建立对鲁棒性的控制。比如AIDS⁸病中，HIV⁹感染CD4-阳性T细胞，当细胞启动抗毒响应时则大量复制。HIV充分利用了T细胞的鲁棒免疫响应机制。有一种治疗策略是：引入一种HIV-1¹⁰携带体crHIV-1^[18]包含顺式而不是反式的病毒组装所需的成分，这种携带体携带了抑制野生HIV-1功能的抗病毒基因。如果成功的话，crHIV-1可能重新建立对系统的控制，使HIV-1病毒进入延长潜伏期状态。

鉴于鲁棒性的重要性，亟需建立一个坚实的数学基础之上的统一的生物鲁棒性理论。首先是建立鲁棒性量化指标。这样的指标应当能够很好地契合实验测量，能够用实验检验系统鲁棒程度。尽管在控制理论中已经有了鲁棒度的定义，并且已经有人开始尝试将其应用到生物系统中^[19]，但是由于生物系统内在的多维性（enormous degree of dimensionality）和非线性，测量所有维数上的干扰是不切实际的。应当有一个符合实际的、理论上的可靠指标，可以用来指导选择少数将被干扰的参数，从而可以测量系统的响应。一些可能有利于生物系统鲁棒性分析的数学方法正在发展^[20]，将推动鲁

棒性的数学量化描述。其次，即使是在控制理论中，也没有一个能够全部涵盖鲁棒性各个方面的统一的理论描述。控制理论经常用于解释包含反馈调节的鲁棒性，但是这只涵盖了鲁棒性的一个方面。控制理论假设系统有一个由设计者制定的系统状态将要达到的点，而在生物系统中显然没有这样的设计者，这样的点是系统平衡态时固有的，而且常常动态变化。因此需要建立一种理论用于描述这种自我决定的发展。

计算机科学在系统生物学中的应用

计算机科学的应用贯穿整个系统生物学研究的始终，这一点在图2中表现较为明显。从信息的获取开始，如何系统、全面、高效、精确、大量地采集数据，如何将这些信息整理成通用的数据格式，如何将这些信息在研究者中传播、共享和使用，如何利用这些信息在计算机上建立生物系统模型，开发各种软件进行模拟、分析和可视化等，这些问题都对计算机科学家提出了挑战。

计算机模拟

如前所述，生物系统组分众多，网络行为极其复杂，不可能直观理解系统的行为。如果不建立系统的模型，将难以对其设计系统性的干扰实验。如果直接进行在体实验，不仅成本相对较高，效率相对较低，而且实际上大量的状态是不可控或不可观的，实验所得数据有限。此外，直接的在体实验也不大可能研究系统在极限状态下的行为。

因此，对生物系统在各个层次建模、仿

⁷ 神经母细胞瘤中多药耐药

⁸ Acquired Immune Deficiency Syndrome，爱滋病，获得性免疫功能丧失综合症

⁹ human immunodeficiency virus，人体免疫缺损病毒，艾滋病病毒

¹⁰ HIV病毒的一种亚型

真、分析和设计，是获得系统性理解的重要手段，也是系统生物学目前的主要工作。在计算机上重建生物体，也是系统生物学的目标之一。这样的重建体应当能够重现生物体对内部和外部的各种响应，能够重现生物体各种涌现特征和行为。这样的模型对于疾病预测、药物开发以及生命本质的解读至关重要。

系统辨识其实就是建模过程，在这个过程中，结合对系统的结构和功能的初步理解，选择适当的建模方式，如微分方程、随机模型、元胞自动机和布尔网络等，方式的选择也将对辨识提出要求。比如，在细胞内信号转导建模中，可以选择不考虑物质浓度和反应在空间上的不均匀性，这时只需要常微分方程组模型；而当需要考虑这些因素时，则需要考虑偏微分方程，需要的工作量就大大增加了，因为目前我们对于物质和反应在细胞内的精确定位尚无突破。模型的选择依赖于研究者对系统结构和功能的理解，以及所研究的内容，此外还需考

虑所需计算资源能否满足等问题。

建立初步模型后，需要通过反复的仿真并与实验结果对照，不断改进模型，以达到二者的最佳匹配。计算机仿真在工程领域是系统设计必不可少的步骤，通过仿真分析理解系统的基本动态特性十分必要而且有效，因此需要开发实用、精确、用户友好的仿真系统^[21]。仿真系统需要与其他工具相结合，如实验数据库、参数优化软件、系统分析软件、实验专家系统以及数据可视化软件等。这些软件及其协同工作如图3和图4^[7]所示。这些软件尤其需要考虑的问题是用户友好性，能够方便生物工作者使用。

生物系统具有极大的复杂性，与一般工程系统有很大区别。寻找新的适合于生物系统的建模理论和描述方法，是亟待解决的问题。现有的模型都不同程度地对所研究的系统作了大量假设，虽然好的模型不一定最忠于实物，但是对于一个我们知之甚少的复杂系统，太多性质描述的缺失，其失真程度有多大是难以预料的。如果现有的各种模型与生物系统在本质上不相容，那么最终这些方法都不能揭示生命的本质。因此，迫切需要从更多不同的角度去看待生物体，寻找最合适的描述方法。

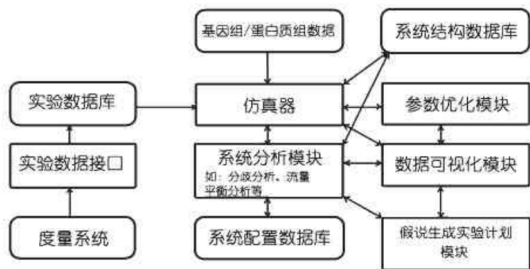


图3 软件工具的关系

通用数据标准与通用软件环境

对生物系统建模需要的大量数据，一般来源于大规模的专用数据库和已发表的文献，而这些数据的格式多种多样，对重复利用带来了许多不便。比如，目前大约有170个通路数据库^[22]，包括代谢、信号转导、蛋白质相互作用和基因表达等，对于研究者来说是非常丰富的资源。但是这些数据库大多拥有各自独特的数据模式、访问方式和文件格式，且不同程度上存在语义差别，因此，数据的使用者不得不不断地适应各个数据库，才能获取所需要的某些特定信息，如果需要综合利用不同数据库提供的资源，则需要更多的努力。对于非专业的使

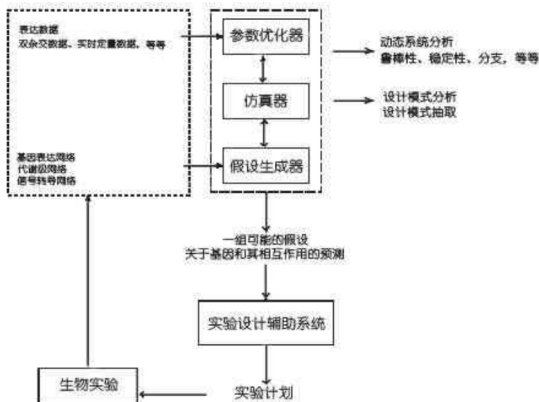


图4 软件工作流程

用者来说,繁多的数据和文件格式也令人望而却步。因此,迫切需要一种通用的数据标准,能够在各种数据库中广泛使用,或者能够作为中间体,实现并简化不同数据格式的转换。这将极大地方便数据的使用者,促进各类研究的进展。同时,数据的提交也更为方便,数据库的容量也因此加速扩大。而大容量的数据库,其包含的信息一般更为全面,对于研究的价值也越大,比如一些预测相互作用的算法,预测结果的准确性和完备性很大程度上取决于所利用的已知的相互作用关系的数目。

目前已经有不少团体致力于这样一些通用数据标准的研究,并且产生了像CellML^{[1][23]}和SBML^{[2][24]}这样的数据标准。这两者均用于不同模拟软件之间的数据交换,希望成为计算机模型数据交换格式的学术标准。这两个标准在不同方面得到了广泛应用(CellML:^[25,26], SBML:^[27,28]),但是,对于众多内容和格式不一的数据库而言,它们的覆盖范围仍然有限,缺乏对一些基本的概念的描述,如“transport”和“RNA¹³”,因此,虽然在软件数据交换上取得了很好的效果,还需要不断完善,才能在数据库中广泛应用。

结语

系统生物学是一个新兴的跨学科交叉领域,综合了生物学、高通量测量技术、系统科学与控制论、计算机科学等诸多领域,需要发展大量新的理解生物系统的思路、实验方法、分析技术和软件工具,需要不同学科工作者的共同努力,融合和应用各个学科,实现对生物体的系统性理解。系统生物学将有助于生物发展、进化机理的研究,有助于对生物现象及其本质的研究,有助于对生命本质的理解。并将极大促进相关学科发展。



孙之荣

清华大学生物系教授,博士生导师,现任生物科学与技术系生物信息学与系统生物学研究室负责人,兼任教育部生物信息学重点实验室副主任、学术秘书。主要研究领域:基因组学、蛋白质组学与生物信息学和系统生物学。目前正围绕功能基因组信息学开展基础和应用研究。

参考文献

- [1] Marie E. Csete, John C. Doyle, Reverse Engineering of biological complexity, Science, Vol 295, 1 may, 2002
- [2] Trey ideker et al., integrated genomic and proteomic analyses of a systematically perturbed metabolic network, Science, Vol 292, 4 may, 2001
- [3] Hood L, A personal view of molecular technology and how it has changed biology, JOURNAL OF PROTEOME RESEARCH 1 (5): 399-409 SEP-OCT 2002
- [4] Leroy Hood, Systems biology: Integrating Technology, Biology, and Computation, Mechanisms of ageing and development, 124(2003)9-16
- [5] Pengcheng Fu, Systems biology: New frontier for system scientists and control engineers, 2004 8th International Conference on Control, Automation, Robotics and Vision, Kunming, China, 6-9th Dec. 2004
- [6] Wiener N. Cybernetics or Control and Communication in the Animal and the Machine. MIT press, 1948

¹¹ CellML是作为Physiome Sciences项目组成部分的细胞标记语言,能够描述(亚)细胞模型的结构和数学关系,可以嵌入原数据和文档,便于模型及其部件的重用性。

¹² 系统生物学标记语言

¹³ Ribonucleic Acid, 核糖核酸

參考文獻

- [7] Kitano H., *Systems Biology: Toward system-level understanding of biological systems*, in *Foundations of Systems Biology*, ed. Hiroaki Kitano, MIT Press, Cambridge, Massachusetts, 2001
- [8] James R. Heath., Michael E. Phelps, Leroy Hood, *NanoSystems Biology, Molecular Imaging and Biology Vol. 5, No. 5*, 312–325. 2003
- [9] Wikswo, J.P. Instrumentation challenges for systems biology, *Sensors*, 2004. Proceedings of IEEE
- [10] <http://www.genome.ad.jp/kegg>
- [11] Dirk Repsilber, Hans Liljenstrom, Siv G.E. Andersson, Reverse engineering of regulatory networks: simulation studies on a genetic algorithm approach for ranking hypotheses, *BioSystems* 66 (2002) 31_41
- [12] Jorg Stelling, Steffen Klamt, Katja Bettenbrock, Stefan Schuster, Ernst Dieter Gilles, Metabolic network structure determines key aspects of functionality and regulation, *NATURE*, VOL 420, 14 NOVEMBER 2002
- [13] Zhike Zi, Kwang-Hyun Cho, Myong-Hee Sung, Xuefeng Xia, Jiashun Zheng, Zhirong Sun, In silico identification of the key components and steps in IFN- γ induced JAK-STAT signaling pathway, *FEBS Letters* 579 (2005) 1101_1108
- [14] Upinder S. Bhalla and Ravi Iyengar*, Emergent Properties of Networks of Biological Signaling Pathways, *SCIENCE* VOL 283 15 JANUARY 1999
- [15] Jacques Ricard, reduction, integration and emergence in biochemical networks, *Biology of the Cell* 96 (2004) 719–725
- [16] Kitano.H, Biological Robustness, *Nature* Nov. 2004 VOL 5
- [17] Kitano, H. et al. Metabolic syndrome and robustness trade-offs. *Diabetes* 53, (Suppl. 3), 1–10 (2004)
- [18] Weinberger, L. S., Schaffer, D. V. & Arkin, A. P. Theoretical design of a gene therapy to prevent AIDS but not human immunodeficiency virus type 1 infection. *J. Virol.* 77, 10028–10036 (2003)
- [19] Ma, L., Iglesias, P. A. Quantifying robustness of biochemical network models. *BMC bioinformatics* 3, 38 (2002)
- [20] Prajna, S. Papachristodoulou, A. in *Proceedings of American Control Conference 2779–2784* (IEEE, Denver, Colorado, 2003)
- [21] Sean Ekins, Yuri Nikolsky and Tatiana Nikolskaya, Techniques: Application of systems biology to absorption, distribution, metabolism, excretion and toxicity
- [22] Michael P. Cary, Gary D. Bader, Chris Sander, Pathway information for systems biology, *EBS Letters* 579 (2005) 1815_1820
- [23] Lloyd, C.M., Halstead, M.D. and Nielsen, P.F. (2004) CellML: its future, present and past. *Prog. Biophys. Mol. Biol.* 85 (2_3), 433_450
- [24] Hucka M, Finney A, Sauro HM, et al. The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models, *BIOINFORMATICS* 19 (4): 524–531 MAR 1 2003
- [25] Garry A, Kohl P, Noble D, Cellular Open Resource (COR): A public CellML based environment for modeling biological function, *INTERNATIONAL JOURNAL OF BIFURCATION AND CHAOS* 13 (12): 3579–3590 DEC 2003
- [26] Nickerson DP, Hunter PJ, Cardiac electro-mechanics: from CellML to the whole heart, *JOURNAL OF MOLECULAR AND CELLULAR CARDIOLOGY* 37 (1): 175–175 JUL 2004
- [27] Kitano H, Funahashi A, Matsuoka Y, et al. Using process diagrams for the graphical representation of biological networks, *NATURE BIOTECHNOLOGY* 23 (8): 961–966 AUG 2005
- [28] Chickarmane V, Paladugu SR, Bergmann F, et al., Bifurcation discovery tool, *BIOINFORMATICS* 21 (18): 3688–3690 SEP 15 2005