电子细胞 Analog-Cell 的并发机制

孙吉贵[1,2], 行 荣[1,2], 卢欣华[1,2], 韩霄松[1,2]
(1. 吉林大学计算机科学与技术学院, 吉林长春 130012; 2. 吉林大学符号计算与知识工程教育部重点实验室, 吉林长春 130012)

摘 要: 结合并发约束理论提出了并发 Analog-Cell 模型框架; 基于扩展的并发约束理论, 提出了并发 Analog-Cell 中定量的随机控制机制以及模拟细胞中 RNA 聚合酶介导转录起始过程的算法模型, 可有效地避免系统中随机错误的发生并提高运行效率; 最后, 提出了并发 Analog-Cell 模型中 DNA 转录过程的模拟算法.

关键词: 电子细胞; 并发约束程序; 生物建模
中图分类号: Q811.4 文献标识码: A 文章编号: 0372-2112 (2009) 04-0786-06

Framework of Concurrent Analog-Cell

SUN Ji-gui[1,2], XING Rong[1,2], LU Xin-hua[1,2], HAN Xiao-song[1,2]
(1. *College of Computer Science and Technology, Jilin University, Changchun, Jilin* 130012, *China*;
2. *Key Laboratory of Symbolic Computation and Knowledge Engineering of Ministry of Education, Jilin University, Changchun, Jilin* 130012, *China*)

Abstract: Analog-Cell represents gene expression at the molecule level. It provides abundant image information. We carried out investigations on concurrent semantic analysis of Analog-Cell in this paper. We introduced concurrent constraint theory into Analog-Cell and presented a framework of concurrent electronic cell model. We designed a quantitative stochastic control mechanism, which was based on the study of stochastic concurrent constraint programming. It is capable of avoiding the random system mistakes and improving the performance efficiency of Analog-Cell. Based on non-deterministic temporal concurrent constraint theory, we presented an algorithm model to simulate the initiation of transcription. Last but not least, we proposed an algorithm to simulate whole transcription process.

Key words: electronic cell; concurrent constraint programming; biological modeling

1 引言

电子细胞在计算机上模拟细胞的功能结构、物质组成以及生命活动的动力学行为和生命现象. 用虚拟现实的方式实现友好人机交互, 以便研究者构造细胞结构和其内外部环境的物质组成, 考察、记录细胞实验现象和功能, 再现细胞生命活动和发现新的生物学现象规律[1]. 采用电子细胞辅助真实细胞实验研究, 设定虚拟细胞环境, 代替和辅助传统医学生物学实验, 重现细胞生命活动的重要环节; 把每个时刻特定位置上特定物质的变化, 通过计算机屏幕或以数据报告的形式提供给研究者. 电子细胞是一个多学科交叉的崭新领域, 除了充分利用现代细胞分子生物学的基础和实验研究成果以外, 还必须依赖物理、数学建模理论和高性能计算领域的研究成果. 数理科学、工程技术与生命科学的共同发展, 引领生命科学研究演化为运用高通量实验技术获取海量生物信息, 并在这些生物信息基础上建立物理、数学模型, 最终通过建模与实验相接合的研究手段来定量阐述生命现象的本质规律的定量研究方法[2]. 电子细胞通常采用动态的、混合的、离散的建模方法, 将半定量模型和定性模型相结合构建出综合的物理、数学模型.

目前有三种较成熟的电子细胞模型. 最早的是日本 Keio 大学 Tomita 教授领导的研究小组所建立的原核细胞能量代谢、细胞膜物质运输模型 E-CELL[3]. 另一个较成熟的模型是美国康涅狄格州州立大学 Leslie MLoew 和 James C Schaff 等学者所建立的真核细胞钙转运模型 Virtual Cell[4], 通过简单的图形模拟钙离子流的移动过程. 第三个较成熟的模型是由美国印第安那大学化学系的 Peter J Ortoleva 教授等人开发的模拟细胞对外界刺激反应、物质运输及基因组变化的虚拟细胞模型 Cyber-

收稿日期: 2008-01-29; 修回日期: 2008-09-24
基金项目: 国家自然科学基金重大项目基金(No. 60496320, 60496321); 国家自然科学基金(No. 60773097, 60873148); 新世纪优秀人才支持计划项目基金; 吉林省科技发展计划项目基金(No. 20060532, 20080107); 欧盟项目基金 TH/Asia Link/010(No. 111084); 吉林省杰出青年基金(No. 20080107, 20080617)

© 1994-2010 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

Cell[5]. 这三种电子细胞模型均存在一个不足: 图形显示方面较为简单、不能满足充分观察细胞内生物化学反应过程的需要. 我们已经实现了国内第一个电子细胞模型 Analog-Cell[6,7].

生物系统模型通常复杂度较高且规模庞大, 用 C++ 等传统的高级程序设计语言仿真实现复杂系统的数学模型遇到了很多软件工程以及软件开发技术方面的难题. 我们需要一种新的语言自然、清晰、简洁地描述复杂模型. 并发约束程序(Concurrent Constraint Programming, CCP)是一种基于约束系统的、并发的多代理计算模型. 扩展的并发约束程序包括: 随机并发约束程序(Stochastic Concurrent Constraint Programming, SCCP)[8,9]、非确定时控并发约束程序(Norr deterministic Temporal Concurrent Cornstraint, NTCC)[10]、混合并发约束程序(Hybrid Concurrent Constraint, HCC)[11,12], 以及随机的非确定时控并发约束(Stochastic Norr determinr istic Temporal Concurrent Constraint, SNTCC)[13]等.

本文利用四种扩展的并发约束程序语言建立生物过程模型, 将并发约束理论和扩展并发约束程序的特点与 Analog-Cell 相结合提出了具有随机性, 时控性以及混合计算特性的并发电子细胞模型框架. 同时, 提出了基于并发机制的 Analog-Cell 中核心反应规则的算法.

本文第二节分别阐述如何利用扩展并发约束语言描述生物模型以及 Analog-Cell 中相关计算机制的设计; 第三节提出了并发电子细胞 Analog-Cell 系统框架以及并发 Analog-Cell 模型中 DNA 转录过程的模拟算法.

2　基于并发约束的电子细胞建模

在构建电子细胞时, 将细胞环境看作一个开放的系统, 系统中每时每刻都在发生着许多生物化学反应, 这些反应共同改变着细胞环境. 我们利用计算机的存储器构建一个"细胞环境", 将生化反应描述为过程, 当环境触发了反应过程, 该过程的调用执行就会改变共享的"细胞环境", 过程之间通过共享的"细胞环境"实现协调和通信.

CCP 是多代理计算模型, 由多个代理和一个共享的约束存储器组成. 关于 CCP 的语义描述和更多深入的细节讨论见参考文献[8, 14].

2.1　并发 Analog-Cell 结合反应随机性控制

SCCP 在基本 CCP 的基础上, 赋予读写操作一个随机系数用于表示更加复杂的动力学特性, 从而支持更为复杂的网络结构描述. SCCP 给每个操作附以一个呈指数分布的连续随机变量 Γ 来表示该操作的执行时间. agents 之间通讯的速度根据系统状态呈现动态变化, 且变化具有随机性, 保证了 SCCP 具有高度灵活性. SCCP 进行生物建模时, 系统中的量化实体用流变量[8]来描述, 例如细胞环境中分子浓度和反应时间等, 这些量化的计算元素都可以用流变量来定义. 系统中表示控制变量的逻辑实体, 如酶促聚合反应等生物化学变化用过程表示. 每个生物化学反应都对应着一个处理流变量的过程. 把活动 agents 与不同的系统行为相关联的同时将变量和量化实体相关联, 意味着将他们看作环境的一部分, 这些活动都有其相应的持续时间和特定属性, 前面提到的系统随机性使得模型依然保持动态随机变化的特点. 实际上, 反应速率与物质实体的存在数量是紧密相关的, 我们采用反应系数来描述这种相关的程度. SCCP 语法结构以及语义的描述见参考文献[9].

酶促聚合反应是 Analog-Cell 中的核心反应规则, 酶 E 作为反应的催化剂与反应底物 S 结合为酶-底物聚合体 ES, 该聚合体可能会重新分解为酶和底物, 或者发生内部变化形成该反应的生成物 P 和酶的复合物. 然后复合物分开产生产物 P, 酶又恢复游离状态, 重新参加反应. 表 1 中的反应模型 1 是用 SCCP 建立的酶促反应模型 enz_ reaction.

已有的 Analog-Cell 中, 所有游离态粒子运动的方向和速度都是随机的, DNA 启动子结合位点作用范围内出现 RNA 聚合酶介导转录的起始是一个随机事件. 通过指定结合位点的作用范围控制反应发生的随机机制和生物细胞内部真实反应发生情况一致, 但是, 由于模型的量化程度和机器硬件的限制, 这样的模拟仍然存在不足之处. 简单的随机性控制机制, 导致了系统内部出现非预期随机性错误. 如: 在 DNA 的转录过程中, 两个游离态的核苷酸分子同时与 DNA 中某匹配的核苷酸结合形成了三聚体而不是核苷酸对. 另外, 由于计算机硬件的限制, 系统内部微粒的运动速度远远达不到真实模拟细胞内生化反应的速度要求. 为了提高现有系统中的反应速度, 需要将结合位点的作用范围扩大. 但是, 串行系统中我们通过反应控制容器存储反应队列, 仅仅扩大反应范围又会导致等待队列过长. 这样的变化增加了运行时刻反应之间的时间延迟.

考虑将 Analog-Cell 系统的特点与 SCCP 中的随机机制相结合, 在 Analog-Cell 中为每个反应结合位点指定一个较大的作用范围, 并且每个反应对应地定义一个随机函数. 使得结合位点作用范围和随机函数共同控制反应的发生. 我们可以通过指定随机函数来控制反应发生的速度. 当 RNA 聚合酶出现在结合位点的作用范围内时, 根据随机函数的返回值决定反应的速度. 由于改进的随机机制较原随机机制而言, 更便于实现模型的量化控制, 很大程度上避免了系统中随机性错误的产生. 上述随机性控制机制是本文算法 1 中随机函数定义所采用的关键技术.

© 1994-2010 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net

表 1 主要的反应模型

反应模型 1 enz_reaction
$reaction(k,[R_1,\cdots,R_n],[P_1,\cdots,P_n]):-\ ask_{r_{MA}(k,[R_1,\cdots,R_n])}(\wedge_{i=1}^{n}(R_i>0)).(\|_{i=1}^{n}tell_{\infty}(R_i=R_i-1)\|\|_{j=1}^{m}tell_{\infty}(P_j=P_j+1)).$ $reaction(k,[R_1,\cdots,R_n],[P_1,\cdots,P_n])$ 其中, $r_{MA}(k,[R_1,\cdots,R_n])=k\cdot X_1\cdots X_n$
$enz_reaction(k_1,k_{-1},k_2,S,E,E,ES,P):-reaction(k_1,[S,E],[ES])\|reaction(k_{-1},[ES],[S,E]\|reaction(k_2,[ES],[E,P])$
反应模型 2 Start_transcription
$Start_transcription:-$ $(when\langle RNA\ Pol>0\wedge distant.RNA\ Pol<range\rangle do\ update_state(firstDNA)).$ $(when\langle get_type(firstDNA)=t\wedge get_state(firstDNA)=before_bind\rangle do\ update_state(firstDNA))$
反应模型 3 s_bond
$s_bond\equiv *\ \rho(P_0\|P_1\|P_2\|P_3\|P_4\|P_5\|P_6\|P_7)$ 其中, $\rho=1;\rho_{ij}(i=0,1,2,3;j=0,1,2,3)$ 表示各状态转换的概率函数, 其返回值是小于 1 的正实数: $P_0=when\langle(N>0\wedge DNA>0),\rho_{01}\rangle do(N=N-1\|N\cdot DNA=N\cdot DNA+1)$ $P_1=when\langle(M>0\wedge DNA>0),\rho_{02}\rangle do(M=M-1\|M\cdot DNA=M\cdot DNA+1)$ $P_2=when\langle(M>0\wedge N\cdot DNA>0),\rho_{12}\rangle do(M=M-1\|M\cdot N\cdot DNA=M\cdot N\cdot DNA+1\|N\cdot DNA=N\cdot DNA-1)$ $P_3=when\langle(N>0\wedge M\cdot DNA>0),\rho_{23}\rangle do(N=N-1\|N\cdot M\cdot DNA=N\cdot M\cdot DNA+1\|M\cdot DNA=M\cdot DNA-1)$ $P_4=when\langle(N\cdot DNA>0),\rho_{10}\rangle do(N=N+1\|N\cdot DNA=N\cdot DNA-1)$ $P_5=when\langle(M\cdot DNA>0),\rho_{20}\rangle do(M=M+1\|M\cdot DNA=M\cdot DNA-1)$ $P_6=when\langle(M\cdot N\cdot DNA>0),\rho_{31}\rangle do(M=M+1\|M\cdot N\cdot DNA=M\cdot N\cdot DNA-1\|N\cdot DNA=N\cdot DNA+1)$ $P_7=when\langle(N\cdot M\cdot DNA>0),\rho_{32}\rangle do(N=N+1\|N\cdot M\cdot DNA=N\cdot M\cdot DNA-1\|M\cdot DNA=M\cdot DNA+1)$

2.2 Analog Cell 中的时控机制

NTCC 是基于 Timed CC 的时间进程演算, 将时间划分为等长的单元, 初始时, 程序接受一个环境触发的约束条件作为过程 P 的输入, P 在初始约束存储器的状态下进入当前时间单元执行这个过程, 等待过程到达一个稳定点时, 输出结果约束到共享存储器, 使得存储器中的约束被更新. 同时, 该稳定点决定了程序进入下一个时间单元时需要执行的剩余过程 Q. 关于语法和语义的更多内容, 见参考文献[10]. 在 NTCC 中, 我们可以用 $M=tell(l<pH_{in}<u)$ 描述细胞液中 pH 值的水平, 由于某个时刻细胞液中的 pH 值不能确定, 所以只能用区间来描述. 当然, 也可以用函数来描述这种约束.

Analog Cell 中模拟的转录延伸过程是指 RNA 聚合酶延 DNA 链移动, 顺次合成 RNA 链. DNA 在聚合酶抵达之前解旋, 在其经过之后又恢复为双螺旋结构. 这一生命过程中的时序性是建模的关键.

串行的 Analog Cell 中, 反应函数是顺序加入反应控制容器并串行执行的. 函数会访问该反应中相应的物质的属性信息, 并更新该反应物的属性信息. 这种机制的作用下, 生物体内原本同时发生的反应在该系统中被人为赋予了先后顺序. 针对该问题, 将 NTCC 时控性的实现机制引入 Analog Cell 中.

并发的 Analog Cell 模型将时间划分为等长单元. 转录起始时, 系统接受初始环境触发信号, 把初始环境作为该反应的执行环境, 进入当前时间单元开始运行并更新存储器中相关数据. 下一个时间单元, 依据最新的共享数据检查新的反应的触发条件能否被满足.

表 1 中的反应模型 2 是并发 Analog Cell 模型中 RNA 聚合酶介导转录起始过程的模型. 该模型表示当 RNA 聚合酶出现在 DNA 结合位点的作用范围内时, DNA 序列中的第一个核苷酸的状态被更新为待结合核膜状态; 若 DNA 第一个核苷酸类型为 t(起始密码子第一位), 且状态为结合核膜前的状态, 则该核苷酸与核膜上任意一点结合, 状态更新为已结合核膜状态. 模型中两个条件判断语句之间是顺序执行的关系. 事实上, 后一条语句中的条件表达式中已蕴含了前一条语句的执行结果, 所以两条语句即使按照并发的方式实现, 整个函数的执行结果也可以正确地模拟该生物过程中的时序关系.

2.3 基于 SNTCC 的分子随机结合反应模型

NTCC 虽然定义了时态却没有提供直接用于描述随机特性的机制. SNTCC 在 NTCC 中加入随机执行操作的机制, 融合了 SCCP 和 NTCC 的特点, 具有随机性和时控性.

Analog Cell 中分子的随机结合反应是系统核心反应规则之一. 表 1 中反应模型 3 是利用 SNTCC 建立的 Analog Cell 中分子间的随机结合反应模型 s_bond. 它是 Analog Cell 中结合反应的基本计算模型. 系统中几乎所

© 1994-2010 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

有反应函数的算法都可以在该模型的基础上通过调整某些参数来完成.

表 1 的反应模型 3 中 P_i 表示系统中的生物化学反应过程, 通过这些过程定义实现系统的状态转换. 例如: 第一个反应 P_0 表示系统状态为 0 的前提下, 当 DNA 结合位点作用范围内 N 的个数大于零时, 发生 P_0 反应概率为 ρ_{01}, 即: 系统以概率 ρ_{01}转换至状态 1. 同理, 初始 0 状态满足 P_1 反应条件时, 以概率 ρ_{02}转换至状态 2. 该概率值是由随机函数返回值决定的, 随机函数的定义依赖细胞生物学, 分子生物学, 以及生物化学等理论知识. P_0 反应结束后, 系统中游离态的 N 数量减一变为零, 二聚体 N•DNA 出现在系统环境中, 触发 P_4 和 P_2 反应的发生. 概率的大小上决定了系统演化的方向.

2.4 Analog-Cell 中的混合特征

Bjorn Carlson 等人共同提出了对混合系统进行建模、控制和仿真的程序语言 HCC. 该语言支持用常微分方程描述的约束和代数约束描述的约束, 这两种约束均采用区间方法[15, 16]进行求解. HCC 即可以描述系统状态达到阈值的离散特征又可以描述反应物浓度的连续变化. 关于 HCC 的语义解释以及实现方法参考文献[11, 12].

Analog-Cell 中颗粒的运动属于连续的无规律的分子热运动, 当受到碰撞时, 粒子还会改变其运动速度和运动方向; 同时系统中存在着的粒子间相互作用力, 也是系统中粒子运动方式的主要影响因素. 模拟这些物理特性是 Analog-Cell 的主要工作. 游离态粒子的随机运动属于该系统模型中的连续变化, 每时每刻都在发生, Analog-Cell 中由系统时钟驱动, 每个时间单元内自动更新空间位置. 设计函数对应 HCC 中 always 语句的语法功能, 用该函数调用那些具有时间连续性要求的反应或者位置变换函数. 例如: Analog-Cell 中所有微粒的空间位置更新定义为: always{ recalculate( ); }. recalculate( ) 是微粒空间位置计算函数. 在具体实现时, 需要将 recalculate 函数指针作为 always 的参数. 其实, 不论是 Analog-Cell 中运动方式的模拟, 或者生物化学反应过程的模拟都是既存在着连续性变化又存在着离散变换.

3 并发电子细胞模型

利用扩展并发约束程序可以实现很多更复杂生物系统的建模. 我们将上述几种扩展机制和 Analog-Cell 相结合, 使得 Analog-Cell 成为了一个具有良好的随机控制机制, 能够准确模拟生物过程时序性, 并具有混合计算特征的并发电子细胞模型.

3.1 并发 Analog-Cell 的系统框架

Analog-Cell 选择了细胞内与基因表达相关的物质如 DNA、核苷酸、mRNA、核糖体、tRNA、氨基酸、多肽链及核糖核酸酶等作为该电子细胞的主要物质组成. 该电子细胞模型的组成部分为: 反应物、基因序列和反应规则. Analog-Cell 通过类型和状态定义来区别不同反应物的不同状态; 反应规则是把反应物、基因序列整合到一起的模型的核心, 反应规则分为两个部分: 反应过程与该反应中进行的状态控制. 关于 Analog-Cell 的更多内容可参考文献[7].

基于扩展 CCP 对 Analog-Cell 改进的基本思想是将细胞内部生物活动和生化反应看作信息处理的过程. 表 2 对 Analog-Cell 和 CCP 进行了比较. 并发的 Analog-Cell 模型中的共享数据包括两部分内容: 空间位置和状态属性. 分别将两类数据存储于两个容器: grid 和 reactant container. grid 用于存储所有粒子的空间位置, reactant container 用于存储反应物的类型和状态. 每个生物化学变化都作为一个独立的过程, 各反应之间互斥地修改共享数据, 更新系统中反应物的类型状态. 触发一个生物化学变化需要满足两种类型的约束, 参与反应的物质必须物理位置相邻并且该微粒的状态满足反应的条件. 也就是说, Analog-Cell 中约束分为两类: 空间约束, 即微粒之间相互作用的约束和反应物的类型状态约束. 其中, 空间约束描述的是通过微粒间的作用力对微粒运动速度和方向的影响; 反应物的类型状态约束则主要用于判断系统环境中反应条件是否成熟. 这样的判断通过访问贡享数据才能完成, 访问共享数据的基本操作包括读取和更新.

表 2　Analog-Cell 和 CCP 比较

Analog-Cell	CCP
空间位置存储器和反应物容器	约束存储器
生物化学反应过程	代理
空间位置约束和反应物状态约束	约束
对共享数据的读取	ask
对共享数据的更新	tell
游离态粒子随机运动以及反应发生的随机性控制	随机性
等长时间单元以及状态控制	时控性
反应过程产生的反应物状态变化	离散变换
粒子空间位置变化	连续变换

图 1 给出了 Analog-Cell 的体系结构. 两个圆角矩形分别表示 grid 和 reactant container. 图中仍用 ask 和 tell 表示两个共享数据区的基本读写操作. 在基本读写操作的基础上构建出具体操作层, 包括: 重新计算微粒间作用力及微粒运动速度和方向的函数 reculculate, 微粒的位置转换函数 move_ partical 以及 grid 中标记已占用位置的函数 add_ occupant. 系统最外层是 clock 和 reaction. clock 会触发 grid 类执行 reculculate 函数, 每个时隔内都会自动更新所有微粒的空间位置信息, 这个函数中考虑了系统的空间位置约束, 例如: 微粒之间是否存在连

© 1994-2010 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net

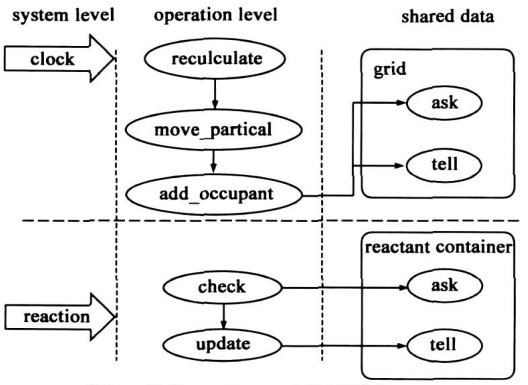


图1 并发Analog-Cell体系结构

接, 是否有互斥力或吸引力的作用, 以及微粒与细胞壁的相互作用力, 这些因素都会影响分子运动的速度和方向. reaction 是 Analog-Cell 系统中添加并发反应函数的操作, 多个反应之间可能并发执行.

3.2 Analog-Cell 中转录过程的模拟

Analog-Cell 模型成功模拟了如下生物过程: 双链 DNA 的解旋; 解旋后, 模板链在 RNA 聚合酶作用下转录得到 mRNA; 在多种酶以及 tRNA 共同作用下, tRNA 携带的氨基酸以 mRNA 为模板生成对应多肽链. Analog-Cell 中模拟转录过程时, 反应中的时序关系就是代码执行的时间关系, 算法执行一次完成一次 DNA 的转录: 顺次将系统中游离的核苷酸合成 mRNA, 算法详见文献[7]. 然而, 实际上细胞内转录起始时, 启动子结合位点并不需要等待上次转录过程完全结束后才可以开始下一次的转录. 只要启动子结合位点空闲, 并且环境中存在反应所需要的酶和所有介导转录起始的物质, 就可以开始下一次转录, 也就是说, 生物细胞内部的转录过程是并行的. 在提出的并发约束框架中, 反应的发生同生物学原理相一致: 当起始位点处满足转录起始条件就可以开始新的转录, 而不必等待上一次转录过程完全结束. 这样使得系统执行效率得到了大幅度提高. 更重要的是, 模拟生命过程更符合生物学的作用原理. 算法 1 描述了模拟真核生物转录的过程.

算法 1 中, 1.1 初始化 DNA 序列中的核苷酸、核膜以及游离核苷酸的状态. 1.2 是描述转录起始过程的 Start_ transcription. 若反应范围内存在 RNA 聚合酶, 则 DNA 第一个核苷酸状态更新为结合核膜前的状态; 若 DNA 第一个核苷酸类型为 t(起始密码子第一位), 且状态为结合核膜前的状态, 则该核苷酸与核膜上任意一点结合, 且状态更新为结合核膜后的状态. 转录起始过程式转录过程中非常重要的一步, 构建 Start_ Transcription 的方法如 2.2 中所述. 1.3 是基于本文 2.3 中的 s_ bond 模型建立的描述转录进行中 RNA 聚合酶、游离态核苷酸和 DNA 序列之间的相互作用. 其中, M 和 N 分别表示 RNA 聚合酶和游离态核苷酸. 依据生物学中的基本原理, 为各转换指定随机函数, 对于可能性极低的转换, 指定其概率 ρ 为 0. 1.4 描述的是 mRNA 与 DNA 的分裂过程. 最后一个核苷酸配对完成后, 1.4 所描述的过程就会被系统调用. 当 mRNA 与 DNA 分裂时, 整条核苷酸链中连接各核苷酸的化学键的断裂几乎是同时发生的, 因此, 核苷酸对的断裂应当并发的发生. 算法 1 中符号 $\|_{forall}$ 表示所有核苷酸对断裂的并发操作. 算法中 un_ bond 表示一个核苷酸对的断裂. un_ bond 与 s_ bond 的构建类似, 不做详细阐述. 分裂过程完成后, 系统完成了一次转录并进入新的状态. 当翻译过程所需要的条件成熟时触发对应的过程.

算法 1 Transcription

```
1.1 Initialization.
1.2 Start_ transcription:
always{
(when 〈RNA Pol > 0 ∧ distant. RNA Pol < range〉 do update_ state
      (firstDNA)).
(when 〈get_ type(firstDNA) = t ∧ get_ state(firstDNA) = before_ bind〉 do
      update_ state(firstDNA))
1.3 transcription:
always{
when〈typematch ∧ unbind〉 do{
* ρ(
when 〈(N > 0∧ DNA > 0∧ distant(N) < range), ρ01〉 do(N = N − 1 ‖ N
    • DNA = N• DNA + 1)
‖ when 〈(M > 0∧ DNA > 0∧ distant(M) < range), ρ02〉 do(M = M − 1
    ‖ M• DNA = M• DNA + 1)
‖ when 〈(M > 0∧ N• DNA > 0∧ distant(M) < range), ρ13〉 do(M = M
    − 1 ‖ M• N• DNA = M• N• DNA + 1 ‖ N• DNA = N• DNA − 1)
‖ when 〈(N > 0∧ M• DNA > 0∧ distant(N) < range), ρ23〉 do(N = N −
    1 ‖ N• M• DNA = N• M• DNA + 1 ‖ M• DNA = M• DNA − 1)
‖ when 〈(N• DNA > 0), ρ10〉 do(N = N + 1 ‖ N• DNA = N• DNA − 1)
‖ when 〈(M• DNA > 0), ρ20〉 do(M = M + 1 ‖ M• DNA = M• DNA − 1)
‖ when 〈(M• N• DNA > 0), ρ31〉 do(M = M + 1 ‖ M• N• DNA = M• N•
    DNA − 1 ‖ N• DNA = N• DNA + 1)
‖ when 〈(N• M• DNA > 0), ρ32〉 do(N = N + 1 ‖ N• M• DNA = N• M•
    DNA − 1 ‖ M• DNA = M• DNA + 1)
)//end *
}//end do
}//end always
1.4 split mRNA and DNA sequence:
always{
 ‖forall when〈state = is_ bond〉{unbond(). update_ state()}
}
```

4 结论与展望

文献[7]提出的 Analog-Cell 中反应根据时间先后依次执行. 在每个时间间隔内执行一个位于容器队首的反应函数. 这样的实现机制使得原本同时发生的生物化学反应之间产生了时间差, 使该电子细胞模型的真

© 1994-2010 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

实性和准确性受到了一定的限制. 同时, 按照触发时间串行执行反应函数的程序结构限制了系统的执行效率. 当用户自定义添加反应物时, 反应容器负载变大, 系统的执行效率就会受到限制. 为了解决这个问题, 通过研究 Analog-Cell 的并发语义, 结合并发约束理论提出了并发电子细胞模型框架; 基于随机并发约束理论提出了并发 Analog-Cell 中随机控制机制的定量模型, 可有效地避免系统中随机错误的发生并提高运行效率; 基于非确定时控并发约束理论, 提出模拟细胞中 RNA 聚合酶介导转录起始过程的算法模型. 最后, 提出了并发 Analog-Cell 模型中 DNA 转录过程的模拟算法.

并发约束程序作为一种更高层次的抽象方法, 提高了用计算机模拟生物系统的能力和效率, 降低了复杂系统建模的难度, 用更加贴近自然语言的方法使生物系统的模型结构更加清晰, 进一步实现系统建模和计算机软件实现的分离. 我们今后的工作将围绕该并发电子细胞模型系统架构的实现展开, 并实现 Analog-Cell 中模拟基因调控过程的算法.

参考文献:

[1] 赵明生, 尚彤, 孙冬泳等. 电子细胞的研究现状与展望[J]. 电子学报, 2001, 29(12A): 1740- 1743.
Zhao Ming-sheng, Shang Tong, Sun Dong-mei, et al. The current state and prospect of e-cell[J]. Acta Electronica Sinica, 2001, 29(12A): 1740- 1743. (in Chinese)

[2] 罗若愚, 李亦学. 系统生物学中建模方法的研究现状与展望[J]. 生命科学, 2007, 19(3): 301- 305.

[3] Nobuyoshi Ishii, Martin Robert, Yoichi Nakayama. Toward large scale modeling of the microbial cell for computersimulation[J]. Journal of Biotechnology, 2004, 113(1- 3): 281- 294.

[4] BM Slepchenko, JC Schaff. Quantitative cell biology with the virtual cell[J]. Trends in Cell Biology, 2003, 13(11): 570- 576.

[5] Gordon Broderick, Melania Ru'aini, Eugene Chan. A life-like virtual cell membrane using discrete automata[OL]. http://www.bioinfo.de/isb/2004/05/0016/, 2004- 11- 27/2006- 05- 30.

[6] Xinhua Lu, Jigui Sun, Ying Ren. The regulation of gene expression in e-cell[A]. Proceedings of the Third International Conference on Natural Computation[C]. Washington, DC, USA: IEEE Computer Society, 2007, vol. 3, 507- 514.

[7] 卢欣华, 孙吉贵. Analog-Cell: 一种新的电子细胞图形模型[J]. 电子学报, 2007, 35(1): 49- 53.
Lu Xin-hua, Sun Ji-gui. Analog-cell: a new pattern model of e-cell[J]. Acta Electronica Sinica, 2007, 35(1): 49- 53. (in Chinese)

[8] Luca Bortolussi, Alberto Policriti. Modeling biological systems in stochastic concurrent constraint programming[A]. Constraints[C]. Hingham, MA, USA: Kluwer Academic Publishers, 2008, 13(1- 2): 66- 90.

[9] C Olarte, C Rueda. A stochastic concurrent constraint based framework to model and verify biological systems[J]. CLEI Electronic Journal, 2006, 9(2): 4.

[10] Julian Gutiérrez, Jorge A Pérez, Camilo Rueda, Frank D Valencia. Timed concurrent constraint programming for analysing biological systems[J]. Electronic Notes in Theoretical Computer Science, 2007, 171(2): 117- 137.

[11] Alexander Bockmayr, Arnaud Courtois. Using hybrid concurrent constraint programming to model dynamic biological systems[A]. Lecture Notes in Computer Science[C]. Springer, 2002, 2401: 85- 99.

[12] Björn Carlson, Vineet Gupta. Hybrid cc with interval constraints[A]. Lecture Notes in Computer Science[C]. Springer, 1998, 1386: 80- 95.

[13] C Olarte, C Rueda. Using stochastic ntcc to model biological systems[A]. Proceedings of Latin American Informatics Conference(CLEI)[C]. Cali Colombia, 2005.

[14] Vijay Saraswat. Concurrent Constraint Programming[M]. London, England, Cambridge, MA: MIT Press, 1993.

[15] Corliss G, Kearfott B. Rogorous global search: industrial applications[J]. J Reliable Computing, 1999, (2): 1- 16.

[16] Moore R. Methods and applications of interval analysis[M]. London: Prenice-Hall Inc, 1979.

作者简介:



孙吉贵　男, 1962 年 11 月-2008 年 3 月. 博士、教授、博士生导师、副院长、符号计算与知识工程教育部重点实验室主任. 研究方向: 人工智能、人工生命. 1993 年博士毕业于吉林大学计算机科学系, 1997 年 7 月晋升为教授; 97 年至 98 年在英国 Ulster 大学留学; 原教育部"符号计算与知识工程"重点实验室主任、吉林大学计算机学院副院长, 兼任中国计算机学会理事、教育部高等学校计算机科学与技术专业教学指导分委员会委员.



行　荣　女, 1984 年出生于山西侯马. 2006 年毕业于吉林大学计算机科学与技术学院, 同年九月进入吉林大学符号计算与知识工程教育部重点实验室. 现为硕士研究生, 从事人工生命和智能信息处理方面的有关研究.
E-mail: xingrong1028@hotmail.com

© 1994-2010 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net