

文章编号:1001-9081(2012)01-0228-06

doi:10.3724/SP.J.1087.2012.00228

# 电子细胞中前体 mRNA 加工过程的模拟与研究

王玉贤<sup>1</sup>, 卢欣华<sup>2,3\*</sup>

(1. 广东松山职业技术学院 计算机系, 广东 韶关 512126; 2. 吉林大学 计算机科学与技术学院, 长春 130012;

3. 符号计算与知识工程教育部重点实验室(吉林大学), 长春 130012)

(\*通信作者电子邮箱 luxh@jlu.edu.cn)

**摘要:** 前体 mRNA 的加工是基因表达过程中必需的一个步骤, 是调节基因表达和产生具有生化活性的蛋白质的重要机制, 现有的电子细胞模型对前体 mRNA 的加工过程模拟得较少。自主开发设计的 Analog-Cell 电子细胞模型定义了反应过程中的反应规则, 提出了模拟前体 mRNA 加工过程的多个算法, 真实地再现了基因表达这一细胞内重要的生命活动, 并得到了比较理想的模拟结果。

**关键词:** 电子细胞; Analog-Cell; mRNA; 生物信息学

**中图分类号:** TP391.92    **文献标志码:** A

## Simulation and research on processing of pre-mRNA in electronic cell

WANG Yu-xian<sup>1</sup>, LU Xin-hua<sup>2,3\*</sup>

(1. Department of Computer Science, Guangdong Songshan Polytechnic College, Shaoguan Guangdong 512126, China;

2. College of Computer Science and Technology, Jilin University, Changchun Jilin 130012, China;

3. Key Laboratory of Symbolic Computation and Knowledge Engineering of Ministry of Education (Jilin University), Changchun Jilin 130012, China)

**Abstract:** The processing of pre-mRNA is one necessary step in gene expression, and it is an important mechanism to regulate the procedure of gene expression and produce the proteins which affect the life activities of cells. The existing electronic cell (E-Cell) models seldom relate to the simulation on the processing of pre-mRNA. The paper proposed an E-Cell model, named Analog-Cell, and reappeared realistically the procedure of gene expression through defining rational reaction rules and several algorithms which simulated the processing of pre-mRNA, thus Analog-Cell obtained recommendable simulation results.

**Key words:** Electronic Cell (E-Cell); Analog-Cell; mRNA; bioinformatics

## 0 引言

细胞的基本生命活动涉及到巨大数量的不同分子及不同分子间的相互作用, 使得这些过程变得异常复杂。生物学家需要对细胞的生命活动进行深入的了解, 获得细胞生命活动每一个步骤的详细数据。为此日本、美国、欧洲以及中国的多所大学、研究所相继研制出电子细胞 (Electronic Cell, E-Cell)<sup>[1-3]</sup>。电子细胞亦称虚拟细胞, 是计算机科学和生命科学相结合的一门新兴学科, 它应用计算机科学的原理和技术, 通过数学的计算和分析, 对细胞的结构和功能进行模拟、分析、整合和应用, 以再现真实细胞的生命活动现象, 为生物学家在计算机上进行实验提供了一个软件平台, 并以此来发掘细胞中潜在的、未被发现的生命活动规律。

细胞为了更好地利用基因组中包含的生物学信息, 代表信息单位的各个基因必须协调表达。这种协调表达决定了细胞中蛋白质组的成分, 从而决定了细胞的生化特征<sup>[4]</sup>。基因组表达为蛋白质组的过程主要分为转录、翻译两个阶段。在真核细胞中, 第一个阶段转录的初始产物, 即各种前体 RNA 多数都不具有生物活性, 只有在细胞内进行加工处理后才具有生物活性<sup>[5]</sup>。该加工过程是对初级转录物进行加帽、多聚腺苷酸化 (Poly(A), 即加尾)、RNA 剪接等使之成为成熟的

mRNA (messenger RNA)、rRNA (ribosomal RNA) 或 tRNA (transfer RNA) 的过程。在现有的电子细胞模型中, 对前体 mRNA 的加工过程模拟得较少。本文提出了模拟前体 mRNA 加工过程的多个算法, 并在自主研制的国内第一个电子细胞图形模型 Analog-Cell<sup>[6-8]</sup> 中得到了较为理想的模拟结果。

## 1 前体 mRNA 的加工

真核细胞前体 mRNA 的加工一般要经过 4 个步骤: mRNA 的 5'端加帽、mRNA 的 3'端多聚腺苷酸化(加尾)、内含子剪接以及对某些碱基的修饰即 RNA 编辑等<sup>[4]</sup>。RNA 编辑在真核细胞的前体 mRNA 中较为少见, 且在生物学中该机制进化的原因尚不清楚<sup>[5]</sup>, 所以本文不探讨 RNA 编辑这一加工过程, 只对前三种前体 mRNA 的加工进行模拟和研究。前体 RNA 的加工过程如图 1 所示。

### 1.1 前体 mRNA 的 5'加帽

在转录起始以后, 转录产物前体 mRNA 的加帽反应立即发生, 通过增加 7-甲基鸟嘌呤 (7-Methyl-Guanosine, 7-MG) 帽子结构修饰前体 mRNA 的 5'端, 一般是在前体 mRNA 达到 30 个核苷酸长度以前完成<sup>[4]</sup>。

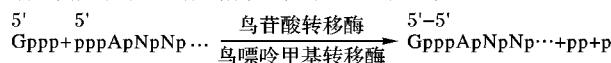
加帽的第一步是加一个附加的鸟苷酸 G 到 RNA 的 5'端, 该反应由鸟苷酸转移酶催化。第二步反应将一个甲基加

收稿日期:2011-05-23;修回日期:2011-07-16。

基金项目:国家自然科学基金资助项目(61003101);欧盟项目基金 TH/Asia Link/010 资助项目(111084);吉林省科技发展计划基金资助项目(20101501);吉林省杰出青年基金资助项目(20080107, 20080617);吉林大学基本科研业务费项目—平台基地建设项目(93K172011704)。

作者简介:王玉贤(1975-), 女, 吉林长春人, 讲师, 硕士, 主要研究方向:计算机建模、人工智能; 卢欣华(1977-), 女, 吉林长春人, 讲师, 博士, 主要研究方向:电子细胞、生物建模。

到鸟嘌呤环的 7 位 N 原子上,使鸟嘌呤变成 7-甲基鸟嘌呤,该反应由鸟嘌呤甲基转移酶催化。加帽完成后,mRNA 上的一个磷酸和鸟苷酸上的两个磷酸被去除,形成一个 5'-5' 键。这种 7-甲基鸟嘌呤结构叫做 0 型帽结构,还有 1 型帽结构和 2 型帽结构,都与 0 型帽结构具有相似的形成方式,它们都是真核细胞中常见的加帽形式。反应过程<sup>[5]</sup>如下:



5' 帽子结构对于 mRNA 翻译为蛋白质具有非常重要的作用,主要表现在以下方面:1)保护 mRNA 5' 端不被降解,5' 帽子可以阻止 mRNA 在翻译过程完成之前被 RNA 酶降解,因为降解是从 5' 端开始的;2)真核生物 mRNA 必须通过 5' 帽子才能被核糖体识别,才能开始翻译过程;3)5' 帽子是 RNA 分子进出细胞核的识别标记,翻译过程必须在细胞核外细胞质中进行;4)5' 帽子涉及第一个内含子剪接复合物的形成,直接影响 mRNA 的剪接效率。

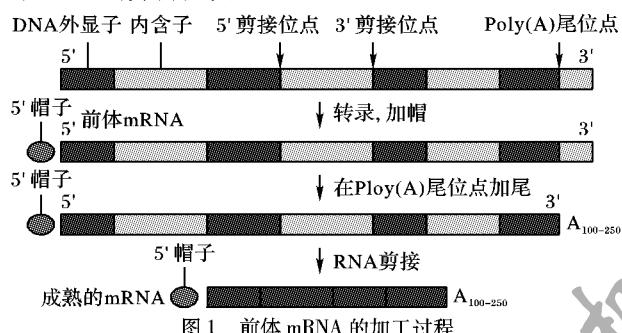


图 1 前体 mRNA 的加工过程

## 1.2 前体 mRNA 的 3' 端多聚腺苷酸化·加尾

所有真核细胞 mRNA 的 3' 端都有一个约 100 到 250 个腺苷酸的 Poly(A) 尾。这些腺苷酸通过 Poly(A) 聚合酶加到前体 mRNA 上,但是并不像 5' 帽子直接作用于 3' 端,而是在其内部的 Poly(A) 尾位点处切断产生一个新的 3' 端,然后加上 Poly(A) 尾,如图 2<sup>[4]</sup>。

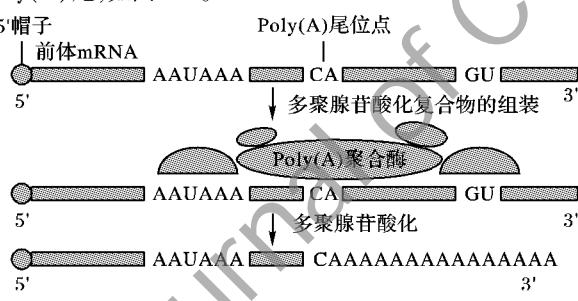


图 2 前体 mRNA 的多聚腺苷酸化

前体 mRNA 的 3' 端含有一个 Poly(A) 信号序列 5'-AAUAAA-3' 作为明显的加尾信号,位于 Poly(A) 尾位点上游 10~30 个核苷酸处。加尾过程首先是 Poly(A) 聚合酶和其他蛋白因子结合于 Poly(A) 信号序列及下游富含 GU 的区域,然后在 Poly(A) 尾位点处切断前体 mRNA,再迅速在新的 3' 末端加上 100~250 个腺苷酸。

对于 Poly(A) 尾的作用生物学家并没有统一的意见,但可能是以下方面:1)增加 mRNA 的稳定性;2)提高 mRNA 的翻译效率;3)Poly(A) 尾影响前体 mRNA 最后一个内含子的剪接,缺少 Poly(A) 使剪接效率降低 5~10 倍<sup>[5]</sup>。

## 1.3 RNA 剪接

真核细胞基因 DNA 中无编码意义而被切除的间插序列称为内含子。被内含子隔开的基因序列,即出现在成熟 RNA

中的序列称为外显子。真核生物前体 mRNA 切除内含子,连接外显子形成成熟的 mRNA 的过程称为 RNA 剪接(RNA Splicing)<sup>[4]</sup>。

一条真核前体 mRNA 可以包含很多内含子,这些内含子必须被切除,然后外显子必须按照正确的顺序连接起来,才能形成一个行使功能的成熟的 mRNA 分子。大多数内含子的两侧边界均有一对保守的序列作为剪接位点,即 5'-GU-3' 和 5'-AG-3',这种称为 GU-AG 的内含子都以相同的方式进行剪接,即由一个蛋白复合物——剪接体催化完成<sup>[9]</sup>。RNA 剪接过程如图 3。

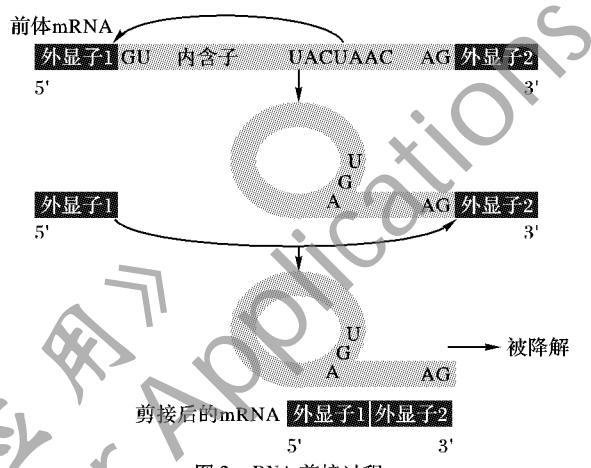


图 3 RNA 剪接过程

RNA 剪接在剪接体中由两步反应完成:第一步,由内含子序列内部的一个保守序列 5'-UACUAAC-3' 中的最后一个 A 向内含子的 5' 端发起反应,导致 5' 端剪接位点断开,同时这个 A 与内含子中的第一个核苷酸 G(即 5'-GU-3' 中的 G)连接,内含子自身成环,形成套索结构;第二步,由内含子上游的外显子 1 的 3' 端向内含子的 3' 端发起反应,导致 3' 端剪接位点断开,套索结构的内含子被释放,然后被降解,同时上游外显子 1 的 3' 端与下游外显子 2 的 5' 端相连,完成剪接过程。

## 2 Analog-Cell 中前体 mRNA 的加工过程模拟

如前所述,真核细胞的转录产物前体 mRNA 都要经过加工过程,包括加到 5' 端的帽子结构,加到 3' 端的 Poly(A) 尾,对内部的内含子进行剪接等,经过加工过程之后前体 mRNA 被加工为成熟的 mRNA,进行基因表达的下一阶段,即翻译过程。Analog-Cell 电子细胞模型<sup>[6~8]</sup>为了更真实、更全面地模拟基因表达过程,加入了加工过程中所必需的调控因子,如 Poly(A) 聚合酶、剪接体复合物等,并制定了合理的加工算法,通过改变反应物的状态值记录反应发生的过程和控制反应发生的条件,因此生动形象地模拟了真核细胞内前体 mRNA 的加工过程,进而完整地模拟了基因表达,为发现真核细胞内基因表达过程的新规律提供了一定的可能性。

### 2.1 Analog-Cell 的反应规则和状态控制机制

根据已知的生物学原理,利用模型确定的建模方法制定出模拟生物化学反应过程的反应规则是电子细胞建模设计的核心。状态控制机制通过对反应过程中反应物状态值的改变,控制反应过程何时进行、何种反应物参与何种反应。

前体 mRNA 的加工过程必须是按照先加帽、再加尾、最后内含子剪接的顺序进行;反应过程又必须是某种特定反应物参与该反应。例如:已加帽或已加尾的 mRNA 状态值应改

变,避免该 mRNA 再次发生加帽或加尾反应。表 1 给出了模型制定的加帽反应的状态定义。

表 1 加帽反应状态定义

反应物	状态值	状态值含义
GTP(g)	0	游离的核糖核苷酸
	1	结合了鸟苷酸转移酶与 mRNA 的核糖核苷酸
	2	结合了鸟嘌呤甲基转移酶与 mRNA 的核糖核苷酸
	17	加帽反应完毕,5'帽子标识
鸟苷酸转移酶	0	游离的鸟苷酸转移酶
	1	与 mRNA 结合后的状态
鸟嘌呤甲基转移酶	0	游离的鸟嘌呤甲基转移酶
	1	与 mRNA 结合后的状态
mRNA	17	转录完毕、未加帽的 mRNA
	18	结合了鸟苷酸转移酶的 mRNA
	19	结合了鸟嘌呤甲基转移酶的 mRNA
	16	加帽反应后的 mRNA

下面分别针对前体 mRNA 的三个加工过程进一步说明 Analog-Cell 的反应规则和状态控制机制。

## 2.2 模拟前体 mRNA 的 5'加帽反应

Analog-Cell 模拟了真核细胞内转录起始之后进行的前体

表 2 前体 mRNA 加帽反应规则

反应底物	底物状态	反应产物	产物状态	反应用到的酶	反应发生条件
前体 mRNA	17	加帽后的 mRNA	16	鸟苷酸转移酶;	转录成功开始; 前体 mRNA 长度未超过 30;
GTP(g)	0	Mra	17	鸟嘌呤甲基转移酶	反应范围内存在 GTP 和所需要的酶

图 4 解释了算法在前体 mRNA 加帽反应中的状态控制机制。状态为 17 的前体 mRNA 的第一个核苷酸 A 在结合 5'帽子后,状态更新为 16,5'帽子 g 的状态更新为 17。

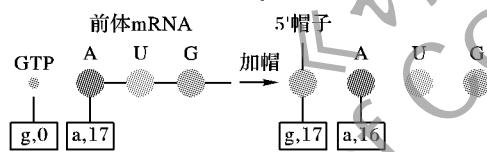


图 4 前体 mRNA 的加帽反应

## 2.3 模拟前体 mRNA 的 3'端加尾反应

Analog-Cell 中模拟前体 mRNA 加尾过程制定的反应规则和反应算法如算法 2 和表 3 所示。

### 算法 2 前体 mRNA 加尾反应算法。

输入 未加尾的前体 mRNA(含位点标识), 腺苷酸 ATP (adenosine-triphosphate), 内切核酸酶和 Poly(A) 聚合酶。

输出 加尾后的 mRNA, 切除掉的一段 mRNA, 内切核酸酶和 Poly(A) 聚合酶。

1) 确认前体 mRNA 3'端上游存在 Poly(A) 信号序列、Poly(A) 尾位点和富含 GU 的区域。

mRNA 加帽过程, 针对前体 mRNA 的加帽反应, 制定的反应规则和反应算法如算法 1 和表 2 所示。

### 算法 1 前体 mRNA 加帽反应算法。

输入 未加帽的前体 mRNA, 鸟苷酸 GTP (guanosine-triphosphate), 鸟苷酸转移酶和鸟嘌呤甲基转移酶。

输出 加帽后的 mRNA, 鸟苷酸转移酶和鸟嘌呤甲基转移酶。

1) 确认前体 mRNA 以 AUG 为开始且第一个核苷酸 A 的状态为转录后、加帽前的状态;

2) if(前体 mRNA 的长度  $\leq 30$ ) &&(5'端反应范围内存在 GTP 及反应所需要的酶)

该 GTP 与前体 mRNA 的第一个核苷酸 A 结合;

3) 更新该 GTP 的状态及前体 mRNA 第一个核苷酸 A 加帽后的状态。

算法 1 的复杂度分析 设前体 mRNA 的长度为  $M$ , 确认前体 mRNA 的状态是否可以触发加帽反应操作的时间复杂度为常数, 记作  $O(1)$ ; GTP 帽子与前体 mRNA 的第一个核苷酸结合时间复杂度为  $O(1)$ ; 更新前体 mRNA 的状态时间复杂度为  $O(1)$ 。因为加帽反应只是前体 mRNA 的第一个核苷酸与一个鸟苷酸 GTP 的合成, 因此, 算法的时间复杂度与前体 mRNA 的长度  $M$  无关。

2) if(前体 mRNA 的 Poly(A) 信号序列和 Poly(A) 尾位点的反应范围内存在 ATP 及反应所需要的酶)

① Poly(A) 聚合酶结合于 Poly(A) 信号序列;

② 内切核酸酶结合于 Poly(A) 尾位点;

③ 切断 Poly(A) 尾位点处的前体 mRNA;

④ 产生一个范围在 100 ~ 250 的随机数  $N$ ;

⑤ while(3'端新加的 ATP 个数 !=  $N$ )

a) 新的 3'端与一个 ATP 结合;

b) 更新该 ATP 和前体 mRNA 的状态。

算法 2 的复杂度分析 同样假设前体 mRNA 的长度为  $M$ , 确认前体 mRNA 的状态是否可以触发加尾反应操作的时间复杂度为常数, 记作  $O(1)$ ; Poly(A) 聚合酶和前体 mRNA 结合时间复杂度为  $O(1)$ ; 内切核酸酶和前体 mRNA 结合时间复杂度为  $O(1)$ ; 断开 Poly(A) 尾位点处的前体 mRNA 时间复杂度为  $O(1)$ ; 在新的 3'端加上  $N$  个腺苷酸 ATP 的时间复杂度为  $O(N)$ ; 更新前体 mRNA 的状态时间复杂度为  $O(1)$ 。因此, 总的算法时间复杂度为  $O(N)$ 。因为加尾反应只是前体 mRNA 的 3'端核苷酸与  $N$  个鸟苷酸 GTP 的合成, 因此, 算法的时间复杂度与前体 mRNA 的长度  $M$  无关。

表 3 前体 mRNA 多聚腺苷酸化反应规则

反应底物	底物状态	反应产物	产物状态	反应用到的酶	反应发生条件
前体 mRNA	17	加尾后的 mRNA;	16	内切核酸酶;	转录成功完成; 前体 mRNA 在细胞核内;
ATP(a)	0	切除掉的一段 mRNA	17	Poly(A) 聚合酶	反应范围内存在 ATP 和反应所需要的酶

图 5 解释了算法在前体 mRNA 加尾反应中的状态控制机制。Poly(A) 尾位点 CA 处的 A 状态为 17, 此处被内切核酸酶切断后, 与反应范围内的 ATP 结合, 状态更新为 16, 而新结合的 ATP 状态更新为 17, 成为新的结合位点。

## 2.4 模拟 RNA 剪接

GU-AG 内含子具有明显的序列特征, 在电子细胞模型的模拟过程中, 剪接体会从前体 mRNA 的 5'端开始移动, 当移动到 5'剪接位点即 GU 的位置时开始剪接反应。但是剪接位

点是类似的,如果一个前体 mRNA 中包含两个以上的内含子,那么很有可能错误的剪接位点被连接。由于到目前为止剪接位点预测技术仍然不是很准确<sup>[10]</sup>,因此在模拟剪接反应时,通过查询核酸序列数据库 GenBank 获取经实验验证过的剪接位点信息,应用于模拟过程。

得到了前体 mRNA 的内含子剪接位点信息之后,Analog-Cell 给前体 mRNA 上的每一个核苷酸定义了一个从 5' 端开始到 3' 端用递增的数值表示的序列属性,表示该核苷酸在序列中的位置,这样也准确地标识出了剪接的位置。剪接反应中的剪接体也有一个初始值为 0 的序列属性,当剪接体与前体 mRNA 结合之后,它的序列属性值会随着它在前体 mRNA 上移动而变化。因此序列属性表示剪接体及核苷酸在前体 mRNA 上的位置,状态值表示参与反应的反应物进行到何种阶段,两个属性一起控制反应发生的条件。

Analog-Cell 中模拟前体 mRNA 内含子剪接过程制定的反

应规则和反应算法如表 4 和算法 3 所示。

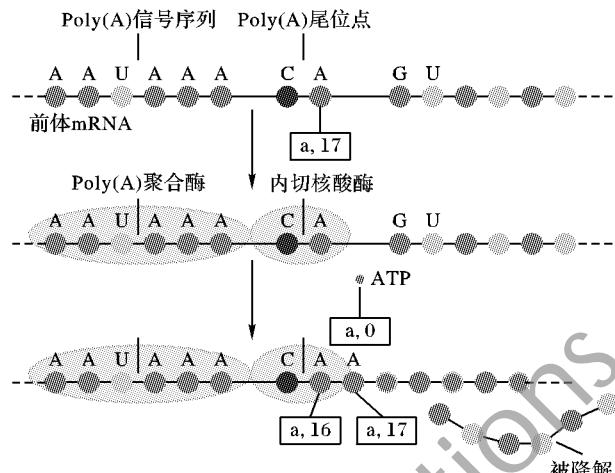


图 5 前体 mRNA 的加尾反应

表 4 前体 mRNA 内含子剪接反应规则

反应底物	底物状态	反应产物	产物状态	反应用到的酶	反应发生条件
前体 mRNA	16	成熟的 mRNA; 剪切掉的套索结构	20, 30	剪接体	转录成功开始; 加帽反应已完成; 反应范围内存在游离的剪接体

### 算法 3 前体 mRNA 内含子剪接反应算法。

输入 未剪接的、已加帽的前体 mRNA, 剪接体。  
输出 成熟的 mRNA, 剪切掉的套索结构, 剪接体。

1) 得到此前体 mRNA 的内含子剪接位点信息, 初始化前体 mRNA、剪接体的序列属性值。

2) 确认前体 mRNA 的每个内含子中都存在保守序列 5'-UACUAAC-3'。

3) 确认前体 mRNA 的第一个核苷酸是 5' 帽子。

4) if(反应范围内存在剪接体)

- ① 剪接体与前体 mRNA 的 5' 帽子结合;
- ② 更新剪接体的状态及序列属性值。

5) while(所有的内含子未去除完毕)

① while(剪接体未移动到内含子 5' 端剪接位点)

a) 剪接体移动一个核苷酸的位置;

b) 更新剪接体的状态及序列属性值。

② 断开 5' 端剪接位点与它前一个核苷酸的连接。

③ 更新剪接体及 5' 端剪接位点的状态。

④ while(剪接体未移动到内含子保守序列 5'-UACUAAC-3' 的最后一个 A 处)

a) 剪接体保持与 5' 剪接位点及 5' 剪接位点前一个核苷酸的连接并移动一个核苷酸的位置;

b) 更新剪接体的状态及序列属性值。

⑤ 5' 端剪接位点与保守序列 5'-UACUAAC-3' 的最后一个 A 结合。

⑥ 断开剪接体与 5' 端剪接位点的连接。

⑦ 更新剪接体、5' 端剪接位点与保守序列 5'-UACUAAC-3' 的最后一个 A 的状态。

⑧ while(剪接体未移动到 3' 端剪接位点)

a) 剪接体与 3' 端剪接位点的下一个核苷酸结合;

b) 断开 3' 端剪接位点与它下一个核苷酸的连接;

c) 与剪接体保持连接的 5' 端剪接位点的前一个核苷酸与 3' 端剪接位点的下一个核苷酸结合;

d) 更新剪接体的状态及序列属性值。

6) 剪接体与 mRNA 脱离, 并恢复为初始状态。

算法 3 的复杂度分析 假设前体 mRNA 的长度为  $M$ , 该前体 mRNA 内共有  $N$  个内含子, 内含子的平均长度为  $K$ , 则初始化前体 mRNA 序列属性的时间复杂度为  $O(M)$ ; 确认前体 mRNA 的状态是否可以触发剪接反应操作的时间复杂度为常数, 记作  $O(1)$ ; 剪接体与前体 mRNA 结合以及更新剪接体的状态和序列属性值时间复杂度都为  $O(1)$ ; 剪去  $N$  个内含子的时间复杂度为  $O(NK)$ ; 剪接体沿着前体 mRNA 的移动反应时间复杂度为  $O(M)$ 。因此, 剪接算法总的时间复杂度为  $O(M)$ 。因为一条前体 mRNA 必定含有一定数量的外显子, 不可能全部都是内含子, 因此  $NK$  一定小于  $M$ 。

图 6 解释了算法在前体 mRNA 内含子剪接反应中的状态控制机制。方框内第一个数值是该反应物的状态, 第二个数值是序列属性值。假设内含子 5' 剪接位点的序列属性值为 4, 3' 剪接位点的序列属性值为 14, 保守序列中最后一个 A 的序列属性值为 11。剪接体移动到 5' 剪接位点后状态更新为 3, 序列属性值更新为 4; 移动到 3' 剪接位点状态更新为 7, 序列属性值更新为 14; 完成剪接后, 脱离 mRNA, 恢复为游离的状态 0; 刚被剪切掉的内含子状态由 16 更新为 20, 整个内含子完全剪切完毕后状态更新为 30, 等待被降解。

### 3 模拟结果

Analog-Cell 电子细胞图形模型在分子水平上模拟了基因表达这一生物化学反应过程。模型运行以后, 一个细胞诞生于窗口内。成功开始转录后, 加帽反应发生; 转录过程结束后, 加尾反应发生(图 7)。

剪接体游离到前体 mRNA 的 5' 帽子处时, 与 5' 帽子结合, 接着移动到内含子的 5' 剪接位点, 断开 5' 剪接位点与前一个核苷酸的连接, 并开始在内含子上移动(图 8)。

剪接体移动到保守序列的最后一个 A 处, 引导内含子的 5' 剪接位点与这个核苷酸 A 建立连接, 形成套索结构(图 9)。

剪接体移动到 3' 剪接位点后, 套索结构的内含子被剪切掉, 最后生成成熟的 mRNA(图 10)。

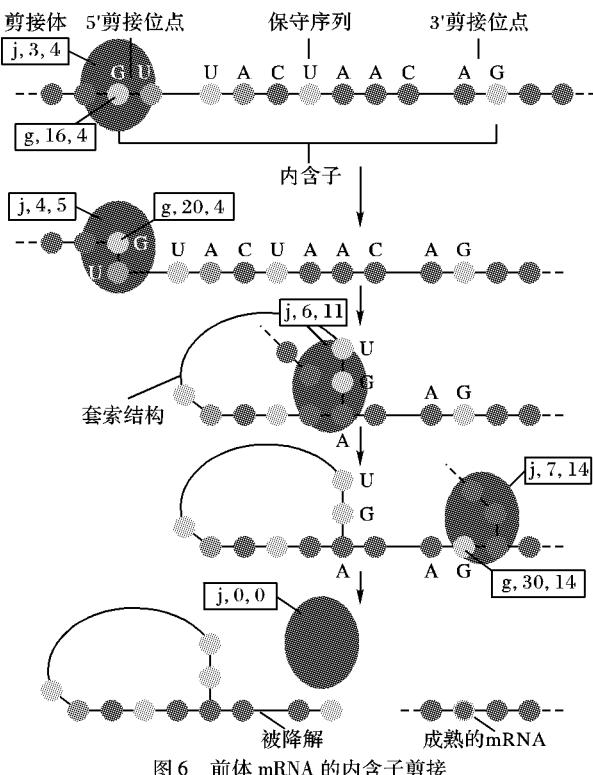


图 6 前体 mRNA 的内含子剪接

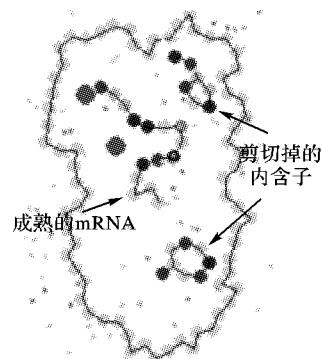


图 10 剪接过程完成

#### 4 相关工作比较

自从 1997 年第一个电子细胞诞生以来,许多国家包括日本、美国都对电子细胞的研究开发进行了大量的投入,使得许多电子细胞模型纷纷涌现出来。但是,由于设计目的的不同,模型之间其特性具有很大的差异。表 5<sup>[11~14]</sup>列出了几种主要的电子细胞之间特性的对比。

表 5 几种主要的电子细胞特性对比

E-cell	mRNA splice	Gene Expression	Eukaryotic Cell	Graphical interface	Nation& Birth time
E-CELL	Yes	Yes	Yes	No	Japan 1998
CyberCell	No	Yes	Yes	Yes	Canada 2003
Virtual Cell	No	No	Yes	No	America 2004
Smart Cell	No	No	No	No	Germany 2004
VICE Cell	No	Yes	No	No	Italy 2005
Analog-Cell	Yes	Yes	Yes	Yes	China 2007

由于生物学界对前体 mRNA 的剪接过程也不甚了解,因此到目前为止关于前体 mRNA 剪接的模拟算法很少。从表 5 可以看出,在较为成熟的电子细胞中,只有 E-CELL 和 Analog-Cell 支持随机动态模拟前体 mRNA 的剪接过程,而 Analog-Cell 还增加了图形化接口。

E-CELL 和 Analog-Cell 的前体 mRNA 加工过程算法都能得到与生物学实验相吻合的数据,图 11 给出了 Analog-Cell 内含子剪接算法的模拟结果与生物学实验数据<sup>[11]</sup>的比较,实验结果包含了基因表达的转录过程,即前体 mRNA 的产生。贴近 X 轴附近的两根曲线代表分别代表生物学实验和 Analog-Cell 模型中刚刚转录完毕、还未发生加帽反应的前体 mRNA;远离 X 轴、mRNA 数量较高的两个曲线同样地也分别代表生物学实验和 Analog-Cell 模型中已发生过加帽反应的前体 mRNA。因为加帽反应在转录开始后就迅速完成,所以未加帽的前体 mRNA 含量在细胞中始终很低。因为生物化学反应的随机性,实际的生物学实验和模型中的模拟结果之间存在细微的误差。

为验证较为复杂的内含子剪接算法的效果,对 E-CELL 的剪接算法和 Analog-Cell 的剪接算法进行对比实验。实验的初始化环境参数如表 6,随着实验中前体 mRNA 的长度增大,反应消耗的能量 GTP 和 ATP、反应需要的催化酶的数量 Enzymes 也需要相应增大。

两种算法得到的两组结果对比如图 12 和 13。实验结果包含了前体 mRNA 的生成即转录过程。两幅图中空心标志

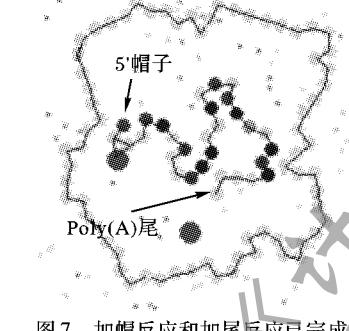


图 7 加帽反应和加尾反应已完成



图 8 剪接体在内含子上移动

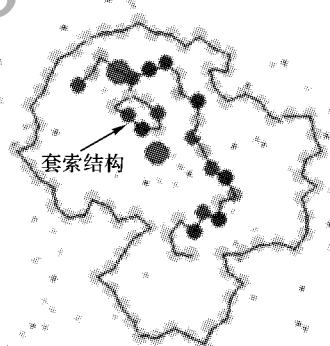


图 9 剪接体引导套索结构形成

代表的曲线都是 Analog-Cell 内含子剪接算法的模拟结果, 实心标志代表的曲线是 E-CELL 的模拟结果, 同样分别地画出了未加帽的前体 mRNA 和已加帽的前体 mRNA 的数量对比。

从两组实验结果的对比可以看出, 无论是第一组数据前体 mRNA 较短、还是第二组数据前体 mRNA 较长的情况下, Analog-Cell 的计算效率都略优于 E-CELL, 完成同样长度的前体 mRNA 剪接过程模拟所花的模拟时间都要略小于 E-CELL。

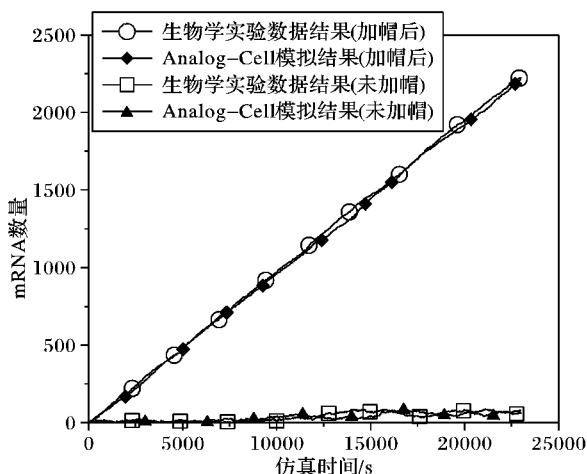


图 11 生物学实验数据与模拟结果的比较

表 6 实验参数表

分组	参数				
	mRNA 长度	内含子个数	GTP	ATP	Enzymes
第一组	235	16	300	300	300
第二组	2261	53	1000	1000	1000

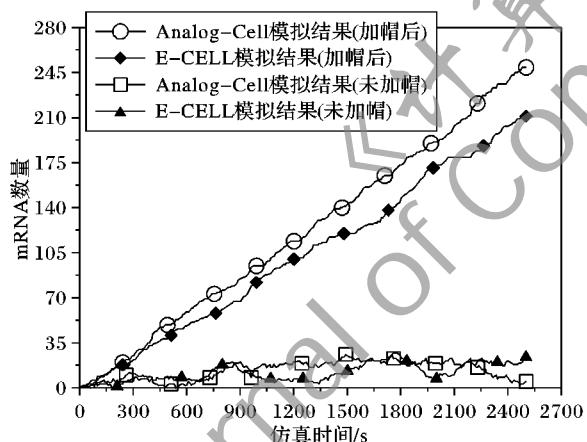


图 12 Analog-Cell 和 E-CELL 模拟结果的比较(第一组)

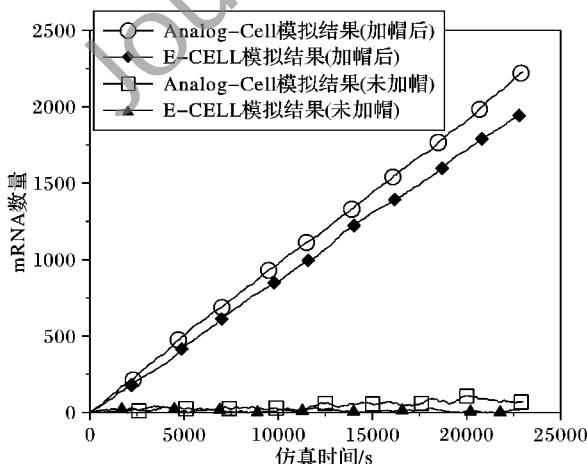


图 13 Analog-Cell 和 E-CELL 模拟结果的比较(第二组)

## 5 结语

电子细胞模型通过在计算机上模拟真实细胞的物质组成、细胞结构和各种生命活动, 向研究者提供了一个可以方便快捷、准确控制反应条件的软件实验平台。基因表达是基因组表达为蛋白质组的重要过程, 其中的转录产物前体 mRNA 必须经过若干的加工过程才能成为成熟的 mRNA, 进而翻译为具有生化活性的蛋白质。

Analog-Cell 是自主建立的国内第一个电子细胞模型<sup>[6]</sup>, 用图形显示方式模拟了真核细胞内与基因表达相关的生物化学过程。本文提出了模拟前体 mRNA 加工过程的 3 个算法, 并在 Analog-Cell 模型中模拟实现, 用户可以直观、形象地观察到整个前体 mRNA 的加工过程, 完善了 Analog-Cell 对于基因表达过程的模拟。模拟结果表明 Analog-Cell 的前体 mRNA 加工算法与生物学的实验数据基本一致, 相对于传统算法, 在不同的实验参数下, 都取得了更优的计算效率, 为进一步模拟真核细胞内其他生命活动、发现生物学的新规律提供了一定的可能性。

### 参考文献:

- MUGGLETON S H. Exceeding human limits [J]. Nature, 2006, 440(7083): 409–410.
- TAKAHASHI K, TANASE-NICOLA S, WOLDE P R. Spatio-temporal correlations can drastically change the response of a MAPK pathway [J]. Proceedings of the National Academy of Sciences of the United States of America, 2010, 107(6): 2473–2478.
- 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.
- BROWN T A. Genomes 3 [M]. New York: Garland Science, 2006.
- WATSON J D, BAKER T A, BELL S P, et al. Molecular biology of the gene [M]. Boston: Pearson Education, 2008: 145–158.
- 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.
- 孙吉贵, 行荣, 卢欣华, 等. 电子细胞 Analog-Cell 的并发机制[J]. 电子学报, 2009, 37(4): 786–791.
- 孙吉贵, 韩霄松, 卢欣华, 等. 真核生物启动子的预测技术[J]. 计算机科学, 2009, 36(1): 5–9, 33.
- NILSEN T W, GRAVELEY B R. Expansion of the eukaryotic proteome by alternative splicing [J]. Nature, 2010, 463(7280): 457–463.
- 刑永强, 张利绒, 罗辽复. 人类基因组盒式外显子和内含子保留的可变剪接位点预测[J]. 生物物理学报, 2008, 24(5): 393–401.
- OHNO H, NAITO Y, NAKAJIMA H, et al. Construction of a biological tissue model based on a single-cell model: A computer simulation of metabolic heterogeneity in the liver lobule [J]. Artificial Life, 2008, 14(1): 3–28.
- ZIMAN A P, WARD C W, RODNEY G G, et al. Quantitative measurement of  $\text{Ca}^{2+}$  in the sarcoplasmic reticulum lumen of mammalian skeletal muscle [J]. Biophysical Journal, 2010, 99(8): 2705–2714.
- DREOSTI E, ODERMATT B, DOROSTKAR M M, et al. A genetically encoded reporter of synaptic activity in vivo [J]. Nature Methods, 2009, 6(12): 883–889.
- PAYNE D F, ORTOLEVA P J. A model for lignin alteration – part I: A kinetic reaction-network model [J]. Organic Geochemistry, 2001, 32(9): 1073–1085.