



(12)发明专利

(10)授权公告号 CN 105760707 B

(45)授权公告日 2018.07.31

(21)申请号 201410805244.7

(22)申请日 2014.12.18

(65)同一申请的已公布的文献号  
申请公布号 CN 105760707 A

(43)申请公布日 2016.07.13

(73)专利权人 中国科学院大连化学物理研究所  
地址 116023 辽宁省大连市中山路457号

(72)发明人 李国辉 张鼎林

(74)专利代理机构 沈阳科苑专利商标代理有限公司 21002

代理人 许宗富

(51)Int.Cl.

G06F 19/12(2011.01)

(56)对比文件

CN 1167151 A,1997.12.10,

CN 101353372 A,2009.01.28,

WO 2012/098505 A1,2012.07.26,

WO 2013/011479 A2,2013.01.24,

Guohui Li等.A Coarse-Grained Normal

Mode Approach for Macromolecules: An Efficient Implementation and Application to Ca<sup>2+</sup>-ATPase.《Biophysical Journal》.2002,第83卷第2457-2474页.

审查员 吕鑫

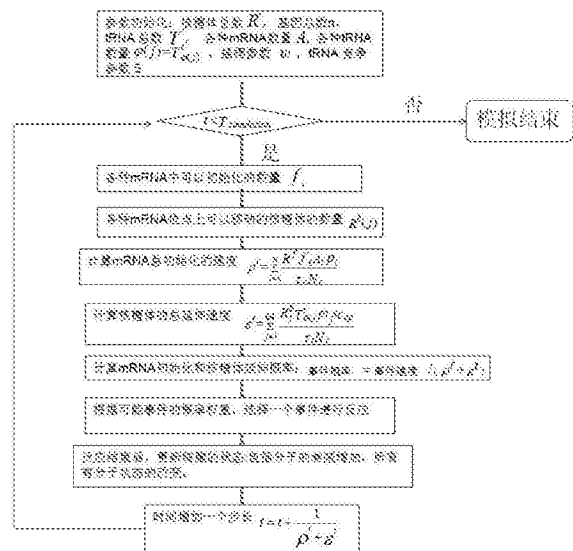
权利要求书1页 说明书11页 附图4页

(54)发明名称

细胞基因翻译过程建模方法

(57)摘要

本发明涉及细胞基因翻译过程建模方法,包括以下步骤:根据跟踪每个mRNA分子的状态并实时更新计数,得到各种mRNA中可以被初始化的数量;根据跟踪每个核糖体分子的状态并实时更新计数,得到各种mRNA位点上可以移动的核糖体的数量;计算mRNA分子的总初始化速度;计算核糖体分子在对应mRNA上的总延伸速度;根据总初始化速度和总延伸速度得到mRNA初始化和核糖体延伸总的概率;根据事件概率的权重,随机选择一个事件进行反应;反应结束后,更新细胞状态;本发明采用基于完全不对称简单排斥过程的原理,同时考虑分子的半衰期,在考虑正确翻译的同时考虑了错误翻译的可能性,使得翻译过程更接近生物问题的本质。



1. 细胞基因翻译过程建模方法,其特征在于包括以下步骤:

1) 根据跟踪每个mRNA分子的状态并实时更新计数,得到各种mRNA中可以被初始化的数量;

2) 根据跟踪每个核糖体分子的状态并实时更新计数,得到各种mRNA位点上可以移动的核糖体的数量;

3) 计算mRNA分子的总初始化速度;

4) 计算核糖体分子在对应mRNA上的总延伸速度;

5) 根据总初始化速度和总延伸速度得到mRNA初始化和核糖体延伸总的概率;

6) 根据事件概率的权重,随机选择一个事件进行反应;

7) 反应结束后,更新细胞状态;并增加一个步长,返回步骤1),直到达到设定时间为止;

所述计算mRNA分子的总初始化速度通过以下公式得到:

$$\rho^t = \sum_{i=1}^n \frac{R^f f_i A_i P_i}{\tau_r N_r}$$

其中, $\rho^t$ 为所有mRNA初始化的速度, $R^f$ 为自由状态的核糖体, $f_i$ 为某种mRNA中可以被初始化的分子数所占的百分比, $A_i$ 为某种基因对应的mRNA的数量, $P_i$ 为某种基因激活的概率, $i$ 为某种基因的id号, $\tau_r$ 为核糖体的运动特征时间, $N_r$ 为核糖体空间位置数量, $n$ 为基因种类数目;

所述计算核糖体分子在对应mRNA上的总延伸速度通过以下公式得到:

$$\varepsilon^t = \sum_{j=1}^{64} \frac{R_j^b T_{\phi(j)}^t \omega_j^s c_{jp}}{\tau_t N_t}$$

其中, $\varepsilon^t$ 为所有核糖体延伸速度, $R_j^b$ 为结合在j类密码子上的核糖体的数量, $T_{\phi(j)}^t$ 为j类密码子对应的tRNA数量, $\omega_j$ 为j类密码子摇摆参数, $s$ 为tRNA竞争参数, $j$ 为某种密码子的id, $\tau_t$ 为tRNA运动特征时间, $N_t$ 为tRNA空间位置数量, $c_{jp}$ 为某种密码子在某种位置上核糖体速度系数;

所述根据总初始化速度和总延伸速度得到mRNA初始化和核糖体延伸总的概率通过以下公式得到:

$$\text{总的概率} = \text{事件速度} / (\rho^t + \varepsilon^t)$$

其中,事件速度为 $\rho^t$ 或 $\varepsilon^t$ , $\rho^t$ 为所有mRNA初始化的速度, $\varepsilon^t$ 为所有核糖体延伸速度。

2. 根据权利要求1所述的细胞基因翻译过程建模方法,其特征在于所述 $c_{jp}$ 通过以下步骤得到:

根据某种密码子在不同位置上核糖体的密度,采用线性拟合的方法对核糖体的密度数据进行线性拟合,得到此类密码子在不同位点上密度,然后再将密度的倒数作为某种密码子在某种位置上核糖体的速度;

将该速度与所有密码子在所有位点上的速度平均值的比值作为某种密码子在某种位置上核糖体速度系数。

## 细胞基因翻译过程建模方法

### 技术领域

[0001] 本发明涉及一种基因翻译过程建模的方法,属于生物信息技术、计算方法与计算机虚拟现实技术。

### 背景技术

[0002] 细胞的新陈代谢、生长和分化等许多基本的生命现象都受到细胞内基因的调控,而基因的这种调控作用则是通过其相应的蛋白质产物来实现的。作为细胞内最基本和最关键的反应之一,越来越多的实验证据表明,蛋白质的翻译对细胞内基因的正常功能发挥起到关键的调控作用。细胞内蛋白质的翻译一般被分为主要的三步:起始、延伸和终止。其中负责基因翻译的关键物质是核糖体:

[0003] 核糖体都由一大一小两个亚基构成,这两个亚基中各具有为合成中的肽链、mRNA、tRNA及翻译因子等分子提供的特殊的结构。核糖体大亚基外侧边缘有三个较明显的凸起,凸起之间有两处凹陷。大亚基内部有一条起源于肽酰转移酶中心附近,横贯大亚基的、为合成中的肽链提供的狭长通道(称为“肽链输出通道”),这条通道长约35aa(约8.5nm)。该通道具有动态的结构,不仅能与新生肽链中的效应模体相互作用调节翻译效率,而且能够调节通过通道的多肽链的共翻译折叠及翻译后修饰。核糖体小亚基大致呈扁弧形,一面也有凹陷。在小亚基一端约三分之一处有一条为mRNA穿过提供的凹槽,该凹槽处的空间局限使得mRNA的解码区呈现U形转角构象。在进行翻译前,两个游离的亚基结合在一起,协同翻译夹在两者之间的mRNA链。核糖体A位点也称“受位”,是核糖体内接受新氨酰-tRNA的位点,主要部分位于大亚基中。在A位处5S rRNA有一段序列能与氨酰-tRNA的T29C环的保守序列互补,便于新氨酰-tRNA移入A位。起始用的tRNA没有该段序列,故它只能直接进入核糖体P位。核糖体P位点也称“供位”,是核糖体内前一个tRNA将其肽基或甲酰甲硫氨酰基转移至后一个tRNA上的位点。该位点在大亚基中的区域含有肽酰转移酶。核糖体E位点,是空载tRNA离开的核糖体的位点。

[0004] 真核细胞基因翻译过程建模方法主要有三种:

[0005] 第一类是基于微分方程组的方法。按照核糖体的结合位点将核糖体分成若干种类——自由核糖体,结合在位点1上的核糖体,结合在位点2上的核糖体,……,结合在位点n上的核糖体,……,最后一个位点上的核糖体。然后针对核糖体的每个状态浓度的变化情况写出化学反应对应的关系,然后针对上步写出来的化学反应关系,得到每个状态浓度变化的微分方程式。这种方法是生物建模最常用的方法之一。因为蛋白质合成由一系列标准的生化反应,这种方法在原则上是可行的。然而,由于伸长的周期性和多个伸长过程可以发生在一个模板,因此这个过程不像大多数其他酶系统那么简单,因此使用单纯的微分方程组建模不容易反应翻译过程的动态的细节——比如多个核糖体在一个mRNA上可能发生相互阻挡。

[0006] 第二类是基于随机方法的,此类方法基于核糖体运动的统计特性。将mRNA看成是格子,核糖体在这些格子上以一定的概率跳跃,主要考虑核糖体在mRNA上移动的跳跃性,特定位点的跳跃概率,相互临近的核糖体之间的空间占位堵塞效应,启动的概率,伸长概率和

终止概率,以上因素决定了mRNA上核糖体的密度。但是以往的模型很少有考虑分子半衰期以及密码错误配对的细节。不能更真实的反应实际的蛋白质翻译过程。

[0007] 第三类是基于TASEP理论的。这类理论被称为“完全不对称的排除过程”或TASEP,是统计方法后续发展,。TASEP的早期版本的特征包括核糖体无限供应的假设,核糖体有统一的沿着mRNA伸长的速率常数,和一个粗粒度的伸长过程,这个过程常被称为跳跃模型。最近修改的TASEP允许分析基于不同密码子的延伸率——不同的密码子有不同的跳跃概率,同时核糖体以及转运rna也不是无限供应的,这种在物质上有限制的模型在本质上更接近真实细胞的生理过程。毕竟真实环境下的物质供应都是有一定的限制的。同第二类一样以往的模型很少有考虑分子半衰期以及密码错误配对的细节,不能更真实的反应实际的蛋白质翻译过程。

[0008] 建模的方法不同但是考虑的问题却是相通的,比如都可以研究蛋白质的合成速度,研究信使RNA密码子的优化,研究在极端条件下,生命体合成蛋白质的特殊状态。不同的建模方法可能会选择不同程度的抽象方式,建模细化程度有差异。有的模型考虑蛋白质分子,核糖体,信使RNA等分子的降解及合成情况,更进一步还可以研究基因表达调控网络。随着实验的深入,相关的参数会越来越多,这为中心法则建模提供了宝贵的依据,使得中心法则建模在理论上更有基础,在应用上更有说服力。

### 发明内容

[0009] 针对上述技术不足,本发明的目的提供一种基于TASEP的建模方法,本方法在传统的TASEP的基础上,考虑了基因翻译过程中出错的可能性。目的是分析整个翻译过程中的速度,考察翻译速度和自由核糖体以及结合状态核糖体数量的变化,考虑各种物质缺乏的状态下,整个翻译过程的状态变化。为真实细胞的基因表达提供理论模型。为实验做理论预测提供依据,为实验做先导,减少实验周期,提高蛋白质合成的效率,降低实验成本。

[0010] 本发明解决其技术问题所采用的技术方案是:细胞基因翻译过程建模方法,包括以下步骤:

[0011] 1) 根据跟踪每个mRNA分子的状态并实时更新计数,得到各种mRNA中可以被初始化的数量;

[0012] 2) 根据跟踪每个核糖体分子的状态并实时更新计数,得到各种mRAN位点上可以移动的核糖体的数量;

[0013] 3) 计算mRNA分子的总初始化速度;

[0014] 4) 计算核糖体分子在对应mRNA上的总延伸速度;

[0015] 5) 根据总初始化速度和总延伸速度得到mRNA初始化和核糖体延伸总的概率;

[0016] 6) 根据事件概率的权重,随机选择一个事件进行反应;

[0017] 7) 反应结束后,更新细胞状态;并增加一个步长,返回步骤1),直到达到设定时间为止。

[0018] 所述计算mRNA分子的总初始化速度通过以下公式得到:

$$[0019] \quad \rho^t = \sum_{i=1}^n \frac{R^f f_i A_i P_i}{\tau_r N_r}$$

[0020] 其中,  $\rho^t$  为所有 mRNA 初始化的速度,  $R^f$  为自由状态的核糖体,  $f_i$  为某种 mRNA 中可以被初始化的分子数所占的百分比,  $A_i$  为某种基因对应的 mRNA 的数量,  $P_i$  为某种基因激活的概率,  $i$  为某种基因的 id 号,  $\tau_r$  为核糖体的运动特征时间,  $N_r$  为核糖体空间位置数量,  $n$  为基因种类数目。

[0021] 所述计算核糖体分子在对应 mRNA 上的总延伸速度通过以下公式得到:

$$[0022] \quad \varepsilon^t = \sum_{j=1}^{64} \frac{R_j^b T_{\phi(j)}^t \omega_j^s c_{jp}}{\tau_t N_t}$$

[0023] 其中,  $\varepsilon^t$  为所有核糖体延伸速度,  $R_j^b$  为结合在  $j$  类密码子上的核糖体的数量,  $T_{\phi(j)}^t$  为  $j$  类密码子对应的 tRNA 数量,  $\omega_j$  为  $j$  类密码子摇摆参数,  $s$  为 tRNA 竞争参数,  $j$  为某种密码子的 id,  $\tau_t$  为 tRNA 运动特征时间,  $N_t$  为 tRNA 空间位置数量,  $c_{jp}$  为某种密码子在某种位置上核糖体速度系数。

[0024] 所述  $c_{jp}$  通过以下步骤得到:

[0025] 根据某种密码子在不同位置上核糖体的密度, 采用线性拟合的方法对核糖体的密度数据进行线性拟合, 得到此类密码子在不同位点上密度, 然后再将密度的倒数作为某种密码子在某种位置上核糖体的速度;

[0026] 将该速度与所有密码子在所有位点上的速度平均值的比值作为某种密码子在某种位置上核糖体速度系数。

[0027] 所述根据总初始化速度和总延伸速度得到 mRNA 初始化和核糖体延伸总的概率通过以下公式得到:

[0028] 总的概率 = 事件速度 / ( $\rho^t + \varepsilon^t$ )

[0029] 其中, 事件速度为  $\rho^t$  或  $\varepsilon^t$ ,  $\rho^t$  为 mRNA 分子的总初始化速度,  $\varepsilon^t$  为核糖体分子在对应 mRNA 上的总延伸速度

[0030] 本发明具有以下有益效果及优点:

[0031] 1. 本发明采用基于完全不对称简单排斥过程的原理, 同时考虑分子的半衰期, 在考虑正确翻译的同时考虑了错误翻译的可能性, 使得翻译过程更接近生物问题的本质, 使用 c++ 语言编程实现。

[0032] 2. 本发明在理论层次上分析核糖体翻译 mRNA 的过程, 在传统的 TASEP 基础上, 尝试引入错误的翻译途径, 更真实的反应了实际的生命过程。

[0033] 3. 在模型中将错误的翻译参数化, 使得本模型可以针对不同的生命环境柔性的变化错误几率, 更真实的反应不同的细胞, 在不同的环境下的蛋白质表达的效果。

[0034] 4. 细胞蛋白质表达过程是细胞生命的核心, 蛋白质表达准确建模可以为电子细胞提供更可靠的性能, 为数字医疗和个性化医疗提交基础。

[0035] 5. 对实验方面, 以及有效的降低实验的盲目性, 提高实验的效率。当前对虚拟细胞的研究尚处于刚刚起步的阶段, 要想应用虚拟细胞对复杂体系进行系统地研究, 还需要搜集大量的实验数据, 针对模拟对象构建更加完备的数据库。在可预测的将来, 从整体上构建细胞的增殖、分化和凋亡模型, 运用复杂系统理论来研究细胞的生命活动规律。

## 附图说明

- [0036] 图1是本发明的方法流程图；
- [0037] 图2是核糖体循环的示意图；
- [0038] 图3是核糖体跳跃模型示意图；
- [0039] 图4是序列不同位置上核糖体平均密度分布图；
- [0040] 图5是在200个位点上每种密码子对应的核糖体密度的线性拟合,得到直线斜率的分布图；
- [0041] 图6是密码子中值密度图；
- [0042] 图7是细胞对抗外部震荡的机制探索图。

## 具体实施方式

- [0043] 下面结合实施例对本发明做进一步的详细说明。
- [0044] 本发明包括以下步骤:基因翻译生物过程的抽象;相关生物问题的数学建模;针对相关的模型用实验数据拟合生化反应参数。
- [0045] 根据基因表达的生物过程进行抽象,得到基因表达的几个关键反应步骤。将核糖体翻译mRNA的过程分成初始化,延伸,结束三个步骤,其中初始化过程做了比较多的简化。整个过程都用核糖体的状态位置来标志,比如说初始化过程就是指核糖体由离散的大小亚基结合在mRNA的启动位点上;此时核糖体的位点标记成为 $R_1^{bound}$ ;以此类推,当延长到位点j的时候此时的核糖体就标志成 $R_j^{bound}$ 。自由核糖体用 $R^{free}$ 表示,自由的时候我们也不将核糖体大小亚基分开看,将二者看成是结合在一起的整体。上面的过程可以用形象的语言表示成小车在(核糖体)一排柱桩上(mRNA)跳跃前进,规定每次只能跳跃一步,而且需要考虑前后核糖体之间的相互阻挡效应,有了上面的抽象的翻译过程,就可以将这个翻译的过程用抽象的数学语言进行建模。
- [0046] 所述建模过程包括以下内容:
- [0047] 模型上采用完全不对称简单排除过程建模,此过程会考虑核糖体之间的相互阻挡。例如选择一个核糖体准备跳跃,如果这个核糖体的下一个位置被另一个核糖体占据了,那么当前的核糖体就没有向下移动的可能,保持原地不动,系统状态不更新。在这个跳跃过程中考虑不同位点跳跃的概率不一样。这主要是基于不同位点的密码相互匹配的t<sub>RNA</sub>的浓度不相同。不同的t<sub>RNA</sub>的浓度导致了跳跃速度的宽慢不一样。同时也考虑了反应过程中错误翻译的可能性,这里将翻译延伸过程中错误概率暂定为0.0001。这个值只是一个建议值,不同的实际情况下这个值会有变化。一旦翻译错误,我们将对错误的情况做简化的处理:如果一个核糖体催化了错误的肽键,那么将直接导致这个肽链的早产即翻译过程提前终止,核糖体在mRNA上释放。
- [0048] 所述参数估计包括以下内容:
- [0049] 在模型参数的估计上,我们既考虑了正确翻译的参数,也考虑了错误翻译的参数。基因表达虽然比较精确,但是各个步骤都有错误的可能。错误翻译的速度和正确翻译的途径不同,速度也不一样。我们根据已有的文献在正确翻译速度的基础上,正确翻译途径和错

误翻译途径反应速度。根据正确翻译的速度估计出了错误翻译的速度。

[0050] 本发明为实现上述目的所采用的技术方案是：基因翻译过程建模方法，包括以下步骤：

[0051] 整理出基因表达的特性，将整个翻译过程分解成关键的几个步骤：

[0052] 在蛋白质合成的时候，翻译组件都进行一系列事件——核糖体大小亚基与mRNA相互结合，在mRNA上延伸——翻译目标mRNA，翻译结束后大小亚基分离。这一系列的聚集和离解的过程被称为核糖体循环。简略地说、翻译始于mRNA，初始tRNA结合到一个自由核糖体小亚基上。小亚基-mRNA-初始tRNA复合体再招募核糖体的大亚基创建一个完整的核糖体，其中mRNA夹在核糖体的大小亚基之间。之后蛋白质的合成开始，翻译的顺序是从mRNA 5'端向3'端延伸。当延伸核糖体遇到一个终止密码子，翻译过程完成，多肽链释放，核糖体大小亚基，mRNA，解离。虽然核糖体的延伸过程只能合成一个多肽，但是每个mRNA上可以同时有多个核糖体移动。

[0053] 翻译起始

[0054] 翻译中的起始过程基本相同：都先由fmet-tRNA、待翻译mRNA和核糖体组成翻译起始复合物(translation initiation complex)。

[0055] 肽链延伸

[0056] 翻译起始后，核糖体沿mRNA链由5'端朝3'端移动。从起始密码子处开始，核糖体利用具有与当前正在翻译的密码子相对应的反密码子的氨酰-tRNA携带的氨基酸合成肽链。核糖体每翻译一个三联体密码子就为延伸中的肽链的C端添加一分子氨基酸，其自身也同时顺着mRNA单向移动一段距离。肽链延伸阶段是一个不断循环进行的过程(直至核糖体到达终止密码子处)。该阶段可再细分为进位、成肽和移位三个步骤。真核及原核翻译延伸阶段的不同主要在于各自延伸因子体系的不同：已发现的原核延伸因子有3种，古菌延伸因子有2种，真核延伸因子有2种。根据核糖体A位内mRNA部分密码子的引导，具有对应反密码子的氨酰-tRNA进入A位的步骤，称为“进位”。原核翻译中，结合了GTP的延伸因子EF-Tu (EF-Tu • GTP) 与氨酰-tRNA形成氨酰-tRNA • EF-Tu • GTP三元复合物并进入核糖体A位。EF-Ts催化水解复合物携带的GTP产生能量完成进位。之后，EF-Tu • GDP脱离核糖体，EF-Tu则释放出GDP并与EF-Ts重新复合形成EF-T，以待再次被利用。真核翻译中，氨酰-tRNA由eEF-1以复合成氨酰-tRNA • eEF-1 • GTP三元复合物的形式带入核糖体的A位。GTP水解后，eEF-1 • GDP离开核糖体。在真菌细胞中，延伸步骤还需要第三种真核延伸因子(eEF-3)参加，其作用是维持翻译的准确性。核糖体的肽酰转移酶催化位于核糖体P位的tRNA上运载的甲酰甲硫氨酰基或肽酰基转移到A位内刚进入核糖体的新氨酰-tRNA转运的氨基酸上，使前者的羧基与后者的氨基之间脱水缩合形成肽键。此步骤称为“转肽”(transpeptidation)。转肽步骤完全由核糖体大亚基内的核酶催化完成。在转肽这一步骤中生成的肽酰-tRNA将占据核糖体A位，而P位中刚卸载甲酰甲硫氨酰基或肽酰基的空载tRNA则将进入核糖体的E位。EF-G将位于核糖体A位的tRNA推入P位的步骤称为“移位”。EF-G具有转位酶活性，能催化GTP水解提供能量，推动处于核糖体A位的肽酰-tRNA，使之进入P位(同时使P位中的tRNA进入E位、并使E位的tRNA排出核糖体)，核糖体也沿mRNA移动一个密码子。核糖体移位后，其A位将再次空出，为新氨酰-tRNA进位提供空间。在肽链延伸的过程中，上述三个步骤是重复循环的。循环每完成一次，肽链的C端便加入一个氨基酸分子，直至翻译进入终止阶段。

## [0057] 翻译终止

[0058] 在核糖体读取至终止密码子时,翻译进入终止阶段。翻译的终止需要释放因子的参与。已发现的原核释放因子有3种,而真核释放因子有2种。原核翻译中,终止密码子由I类释放因子能识别。其中,RF-1能识别终止密码子UAA、UAG;而RF-2能识别UAA、UGA。RF-1或RF-2进入核糖体A位与终止密码子互补配对后使翻译停止。RF-3可使核糖体内部的肽酰转移酶改变构象,从而发挥酯酶活性水解多肽与tRNA之间的化学键。在RR作用下,空载tRNA、mRNA和RF都与核糖体分离,核糖体的大、小亚基也发生解聚。在真核翻译中,三种终止密码子都由eRF-1识别。真核翻译的终止过程与原核翻译的也大体相同。

## [0059] 核糖体再循环

[0060] 核糖体再循环是核糖体参与的翻译中的最后一个过程。在该过程中,核糖体复合物发生解体,以便投入下一次使用。该过程是在核糖体再循环因子(RRF)和参与蛋白质合成过程中转位的延伸因子(EF-G)的协同作用下完成的。核糖体在其再循环过程中由RRF的结构域I识别结合核糖体的功能。RRF的结构域II则具有将核糖体解离为大、小亚基的能力。

[0061] 将上面所述的过程根据核糖体在mRNA上的位置进行分类就是四个状态:初始化,延伸,终止,解散。其中初始化过程相对复杂。在本方法中过程分解使用核糖体的状态和其所处的mRNA上的位置来标志,比如说初始化过程就是指核糖体由离散的大小亚基结合在mRNA的启动位点上;此时核糖体的位点标记成为 $R_1^{bound}$ ;以此类推,当延长到位点j的时候此时的核糖体就标志成 $R_j^{bound}$ 。自由核糖体用 $R^{free}$ 表示。上面的过程可以用形象的语言表示成小车在(核糖体)一排柱桩上(mRNA)跳跃前进,规定每次只能跳跃一步,而且需要考虑前后核糖体之间的相互阻挡的问题,有了上面的抽象的翻译过程,就可以将这个翻译的过程用抽象的数学语言进行建模。如图2所示。

[0062] 如图1所示,细胞基因翻译过程的建模方法如下:

[0063] 初始化参数:包括核糖体总数 $R^t$ ,基因总数 $n$ ,tRNA总数 $T^t$ ,各种mRNA数量 $A_i$ ,各种tRNA数量 $T_{\phi(j)}^t$ ,摇摆参数 $\omega$ ,tRNA竞争参数 $S$ 。

[0064] 判断时间 $t$ 是否小于模拟设定时间 $T_{simulation}$ ;如果小于,则执行下一步,否则模拟结束。

[0065] 根据跟踪每个mRNA分子的状态,更新计数,得到各种mRNA中可以初始化的数量;

“可以初始化”表示mRNA初始结合位点上没有核糖体,数量为 $\sum_{i=1}^n f_i A_i$ , $f_i$ 为某种mRNA中

可以被初始化的分子数所占的百分比, $A_i$ 为某种基因对应的mRNA的数量, $n$ 为基因种类数目。

[0066] 根据跟踪每个核糖体分子的状态,更新计数,得到各种mRNA位点上可以移动的核糖体的数量;“可以移动的核糖体”表示mRNA下一个位点上没有核糖体,数量为 $\sum_{j=1}^{64} R_j^b$ , $R_j^b$

为结合在j类密码子上的核糖体的数量。

[0067] 计算mRNA的初始化速度:

[0068] 计算核糖体在mRNA上的延伸速度:



[0069] 根据上述参数计算mRNA初始化和核糖体延伸的概率:

[0070] 根据事件概率的权重,采用轮盘赌的方法随机选择一个事件( $\rho^t$ 或 $\varepsilon^t$ )进行反应(如:翻译起始和肽链延伸反应);反应结束后,更新细胞状态。考虑分子的半衰期和错误概率的影响,在细胞状态更新时,进行分子衰变,淘汰衰变的分子,然后找出错误配对的核糖体,并结束翻译过程。

[0071] 时间增加一个步长:

$$[0072] \quad \nabla t = \frac{1}{\rho^t + \varepsilon^t}, \quad t = t + \nabla t,$$

返回判断时间t是否小于模拟设定时间 $T_{\text{simulation}}$ 。

[0073] 所述计算mRNA的总初始化速度通过以下公式得到:

$$[0074] \quad \rho^t = \sum_{i=1}^n \frac{R^f f_i A_i P_i}{\tau_r N_r}$$

[0075] 其中, $\rho^t$ 为所有mRNA初始化的速度, $R^f$ 为自由状态的核糖体, $f_i$ 为某种mRNA中可以被初始化的分子数所占的百分比, $A_i$ 为某种基因对应的mRNA的数量, $P_i$ 为某种基因激活的概率, $i$ 为某种基因的id号, $\tau_r$ 为核糖体的运动特征时间, $N_r$ 为核糖体空间位置数量, $n$ 为基因种类数目。

[0076] 所述计算核糖体在mRNA上的总延伸速度通过以下公式

$$[0077] \quad \varepsilon^t = \sum_{j=1}^{64} \frac{R_j^b T_{\phi(j)}^t \omega_j s c_{jp}}{\tau_t N_t}$$

[0078] 其中, $\varepsilon^t$ 为所有核糖体延伸速度, $R_j^b$ 为结合在j类密码子上的核糖体的数量,

$T_{\phi(j)}^t$ 为j类密码子对应的tRNA数量, $\omega_j$ 为j类密码子与tRNA配对的摇摆参数, $s$ 为tRNA竞争参数, $j$ 为某种密码子的id, $\tau_t$ 为tRNA运动特征时间, $N_t$ 为tRNA空间位置数量, $c_{jp}$ 为某种密码子在某种位置上核糖体速度系数。

[0079] 这种延伸速度考虑了真实实验的结果,集中考查了不同密码子的速度差异,尽管导致这种差异的详细因素可能还不太清楚,但是我们对实验数据做了线性拟合,得到了针对不同密码子速度的关键参数 $c_{jp}$ 。

[0080] 所述分子错误机制——每次延伸过程都考虑错误可能,一旦出现匹配错误的tRNA,将导致相关核糖体过早结束延伸过程。

[0081] 所述分子半衰期,每次细胞更新都考虑半衰期的可能,衰变的分子将直接退出翻译过程:核糖体的衰变导致这个核糖体的翻译结束,核糖体退出系统,mRNA的衰变导致这个mRNA上的所有核糖体翻译都变成自由态,相关翻译过程早产,对应的多肽链为失败的多肽链。通过半衰期公式估算指定时间步长的衰变概率:

[0082] 半衰期规律： $\frac{1}{2}N = N e^{-\lambda t_{1/2}}$  其中 $\lambda = \ln 2 / t_{1/2}$ ,  $t_{1/2}$ 为半衰期。

[0083] 每一个分子在指定的时间步长 $t_s$ 衰变的概率为： $p_d = 1 - e^{-\lambda t_s}$

[0084] 模型介绍：模型上采用完全不对称简单排除过程建模——此过程会考虑核糖体之间的相互阻挡。例如选择一个核糖体准备跳跃，如果这个核糖体的下一个位置被另一个核糖体占据了，那么当前的核糖体就没有向下移动的可能——保持原地不动，系统状态不更新。在这个跳跃过程中考虑不同的位点跳跃的概率不一样。这主要是基于不同位点的密码子相互匹配的t\_RNA的浓度不相同。不同的t\_RNA的浓度导致了跳跃速度的宽慢不一样。密码子对应的序号表格如表一：

[0085] 表一

[0086]

1	AAA	17	TAA	33	GAA	49	CAA
2	AAT	18	TAT	34	GAT	50	CAT
3	AAG	19	TAG	35	GAG	51	CAG
4	AAC	20	TAC	36	GAC	52	CAC
5	ATA	21	TTA	37	GTA	53	CTA
6	ATT	22	TTT	38	GTT	54	CTT
7	ATG	23	TTG	39	GTG	55	CTG
8	ATC	24	TTC	40	GTC	56	CTC
9	AGA	25	TGA	41	GGA	57	CGA
10	AGT	26	TGT	42	GGT	58	CGT
11	AGG	27	TGG	43	GGG	59	CGG
12	AGC	28	TGC	44	GGC	60	CGC
13	ACA	29	TCA	45	GCA	61	CCA
14	ACT	30	TCT	46	GCT	62	CCT
15	ACG	31	TCG	47	GCG	63	CCG
16	ACC	32	TCC	48	GCC	64	CCC

[0087] 各种密码子的斜坡数据：图4是AAA密码子的数据，不同密码子的下降速度不同。斜坡的其中一个原因是因为错误的产生，每次前进一步都会产生错误，有一定的错误几率，衰减的规律遵守指数衰减。将这部分进行数学建模。 $P(t) = e^{-\lambda t}$ ,  $P(t)$ 为经过时间 $t$ 后分子不衰变的几率， $\lambda$ 为衰变反应参数。

[0088] 为了提高模拟的速度我们对前200个密码子做线性拟合，200之后的密码子的参数使用第200位置上的参数。如图4所示。

[0089] 针对所有密码子的浓度在不同位点上的中间值做线性拟合，所有密码子的斜率如图5所示。

[0090] 如图3所示，参数估计，使用实验数据估计出模型中参数。不同密码子核糖体浓度和位置的关系：每种密码子密度都有下降的趋势，我们对每种密码子做线性拟合 $y = kx + b$ ,  $k$ 、 $b$ 为系数， $x$ 就是表示序列的位置， $y$ 表示位置上核糖体的密度；得到此类密码子在不同位

点上密度,然后再将密度的倒数作为某种密码子在某种位置上核糖体的速度。序列的位置和位置上核糖体的密度的原始数据参考论文“Composite Effects of Gene Determinants on the Translation Speed and Density of Ribosomes”,作者为Nicholas T.Ingolia USA。对核糖体做线性拟合后的密度速度整理得到下面的参数。

[0091] rd\_mid<sub>j</sub>表示j类密码子在200个位点上的密度取的中值。

[0092] rd\_ave<sub>jp</sub>表示j类密码子在p位点上的密度。

[0093] vcode<sub>j</sub>=1/rd\_mid<sub>j</sub>,vcode<sub>j</sub>代表j类密码子在所有位点上速度。

[0094] vcode<sub>jp</sub>=1/rd\_ave<sub>jp</sub>,vcode<sub>jp</sub>代表j类密码子在p位点上速度。

[0095] median\_rda<sub>11</sub>,表示对所有密码子的平均密度取中值。

[0096] median\_va<sub>11</sub>=1/median\_rda<sub>11</sub>,median\_va<sub>11</sub>代表所有密码子在所有位点上的速度。

[0097] c<sub>jp</sub>=vcode<sub>jp</sub>/median\_va<sub>11</sub>,c<sub>jp</sub>代表j类密码在p位点上的速度与整体速度的比值。

[0098] 数据格式片段如下,其中每一行代表了一种密码子。

	<i>k</i>	<i>b</i>	<i>rd_mid</i>	<i>vcode</i>	<i>C<sub>cp</sub></i>
[0099]	-0.00196	0.67824	0.48237	2.07309	0.85318
	-0.00143	0.51834	0.37541	2.66376	1.09627
	-0.00233	0.89090	0.65759	1.52071	0.62585
	-0.00270	0.97461	0.70443	1.41959	0.58423

[0100] 密码子中值密度如图6所示。

[0101] 对斜坡的一种解释:初始部分衰减的剧烈,中部比较平缓,尾部再次比较距离——这部分可能其他因素起主要作用。

[0102] 同时也考虑了反应过程中错误翻译的可能性,这里将翻译延伸过程中错误概率暂定为0.0001。这个值只是一个建议值,不同的实际情况下这个值会有变化。一旦翻译错误,我们将对错误的情况做简化的处理——如果一个核糖体催化了错误的肽键,那么将直接导致这个肽链的早产——翻译过程提前终止,核糖体在mRNA上释放。实验中已经有的参数如表二:

[0103] 表二

	molecule	number per cell/100000	fixed?
[0104]	eIF2B	0.3	
	eIF2GDP	1.8	
	tRNA	20	yes
	eIF1	2.5	
	eIF3	1	
	eIF5	0.483	
	eIF1A	0.5	
	40S	2	yes
	eIF4E	3.4	
	eIF4G	0.175	
	mRNA	0.15	yes
	PabP	1.98	
	eIF4A	8	
	eIF4B	1.55	
	eIF5B	0.134	
	60S	2	yes

[0105] 其中,molecule表示各种分子,number per cell/100000表示针对一个细胞中每

一个数字后面要乘上100000, fixed表示固定参数。

[0106] 错误翻译情况下参数确定:我们考虑了正确翻译和错误翻译共存的情况,在真实的情况下,正确和错误翻译应该始终伴随。而且有些情况下错误的概率并不低。下表是翻译过程中各个步骤可能发生错误的概率的值。将错误情况放在模型中给相关事件一个错误的几率。本模型中主要考虑延伸过程中的错误事件,当一个错误的延伸产生后,本模型中处理的方式是终止翻译,释放核糖体的大小亚基。根据表三我们将延伸过程错误的概率暂定为0.0001——这个参数可以调整。

[0107] 表三

[0108]

翻译的过程	相关蛋白	底物	识别方式	错误概率
氨基酸激活	aaRS	Aminoacids tRNA	底物选择 转位前编辑 顺式转位后编辑 反式转位后编辑	0.0001-0.00001
aa-tRNA转移	EF-Tu	aa-tRNA	底物选择	?
初始化	IF Ribosome	Initiator aa-tRNA Initiation codon	底物选择	?
延伸	EF Ribosome	aa-tRNA Peptidyl-tRNA	底物选择 校正——质量控制	0.001-0.0001
终止	RF Ribosome	Stop codon	底物选择	0.001-0.000001

[0109] 错误导致有效的蛋白质数量少了,化学反应速度和物质浓度的乘积有关系,导致合成蛋白质速度明显降低,但是还要一个平衡,这就是内部的调整,可以估计出来这种调节的能力。蛋白质合成的速度为何是当前的速度,可以认为当前速度是当前条件下的一个极值点——但是其目标并不清楚,总的来说既应该有一定的速度,同时也要尽量节约能量。速度快和能量节省是不可兼得,而且这个目标可能在不同的时候有不同侧重点,在细胞生长旺盛的生理阶段,当外部物质充足的情况下,速度就是关键因素——这种情况下能量消耗可以做出点让步;在外部物质非常缺乏,同时生长不旺盛的阶段,能量的有效利用就比较关键。

[0110] 细胞对抗外部震荡的机制建模探索。内部调整:主要思想是惰性思想,为变化提供阻力——基本思想是提供一个缓冲区,物质缺乏的时候可以提供少量的物质,物质过于丰富的时候,可以储存物质。氨基酸消耗大概看成蛋白质合成的消耗,可以实时更新,其他分子的消耗就是半衰期,前期可以用一个统一的一个参数,体内物质存储量有一定的量,存储物质的消耗也是按照半衰期的速度前进。这部分的模型初步建立。几个状态参量,外部输出的量相对平衡值的比值 $x$ 。细胞中当前的量相对平衡值的比值 $y_1$ ,物质存储的量相对平衡值的比值 $y_2$ , $y_3$ 物质的流逝。相关示意图如图7所示。

[0111] 其中核糖体成份相对比较复杂,自由的核糖体是不是都是处于大小亚基状态,或者还有更底层的划分,其中的含量如何?各个状态的百分比如何?如果说有更细致的状态存

在,那么自由核糖体的有效浓度会更低。模型目前没有考虑这个细节。

[0112] 半衰期暂定半小时(不同的实际情况有很大的差异,需要根据实际的细胞和状态设定)。想要模拟细胞一定时间的行为,跳跃模型并不方便,但是可以通过跳跃计数来研究,核糖体跳跃时间可以估计出来,真核细胞中核糖体跳跃一次大概0.2秒(不同的细胞这个值有差异),那么我可以记录核糖体在信使上停留的时间,一个核糖体离开后再次标记一个初始核糖体,累积半小时这个核糖体就要面临半衰期事件,蛋白质也一样,核糖体也一样,于是可以用这个模型研究饥饿状态核糖体等物质的分布变化。

[0113] 某种物质缺乏下,细胞的补偿机制,为了保证物质浓度尽量不变,相关反应平衡点要移动。某种氨基酸缺乏可能导致翻译过程这种氨基酸相关位置的翻译错误率升高,通过扩散过来的氨基酸总不是想要的,就会给错误的氨基酸增加了机会,同时也会因此而产生跳跃困难,降低蛋白质合成效率,进而影响生物生长。

[0114] 可以想象正常每个位点的错误概率是0.0001,如果饥饿程度达到0.1.那么错误的几率就可能升高10倍,如果一个多肽链是500个氨基酸,那么正确的肽链只可能是一半。错误的蛋白质如果也和正确的蛋白质掺乎在一起,就麻烦了,会扰乱正常的功能,比如一个核糖体成份众多,其中某些成分不正常,那么直接的结果是这个核糖体可能不正常。如果错误的蛋白质不能马上分解,还会浪费本来已经非常缺乏的物质。这就是恶性循环,这需要实验验证。但是生物的进化显然不会任这种情况发生,错误翻译提前释放就是其中一个措施。

[0115] 由上面的机制来分析一下细胞死亡的必要性,我觉得情况可能如下:我把修复代价称为抓小偷代价。简单解释一下:想在一个房间中抓一个小偷相对容易,但是要在一个国家抓一个小偷则要难得多。这就是破案的代价。类似的问题,细胞中出现错误的可能多种多样,但是相关的修复机制找到这个错误然后把这个错误修复就需要代价。错误多了,修复的代价也非线性的增加,如果修复的代价太大了其实就不如报废了——重新开始,这很可能是细胞适应有限资源的结果。这其中有混沌效应,错误随机产生,产生的错误会影响相关的功能,于是错误的结果再导致新的错误。

[0116] 基本参数解释和部分使用值:

[0117] 核糖体数量 $R^t$ :200000;mRNA数量 $A^t$ :60000;tRNA数量 $T^t$ :3300000;tRNA种类数量 $T_n$ :41;j类tRNA数量 $T_j^t$ :12000-190000;i类mRNA数量 $A_i$ :1-1254;某种基因初始化概率 $P_i$ : $3.5 \times 10^{-6}$ -0.115;基因数量 $n$ :3795;tRNA竞争系数 $S$ : $7.78 \times 10^{-4}$ ;核糖体运动特征时间 $\tau_r$ : $5 \times 10^{-4}$ s;tRNA运动特征时间 $\tau_t$ : $4.45 \times 10^{-7}$ s;运动特征时间(Characteristic time of ribosomes)表示一个指定的分子通过随机运动到达指定位置的期望时间,为设定值。核糖体在一定体积 $V$ 中的位置数量 $N_r$ : $1.56 \times 10^6$ ;tRNA在一定体积 $V$ 中的位置数量 $N_t$ : $1.24 \times 10^7$ ;摇摆系数 $\omega$ :0.64; $c_{jp}$ 代表j类密码在p位点上的速度与整体速度的比值。

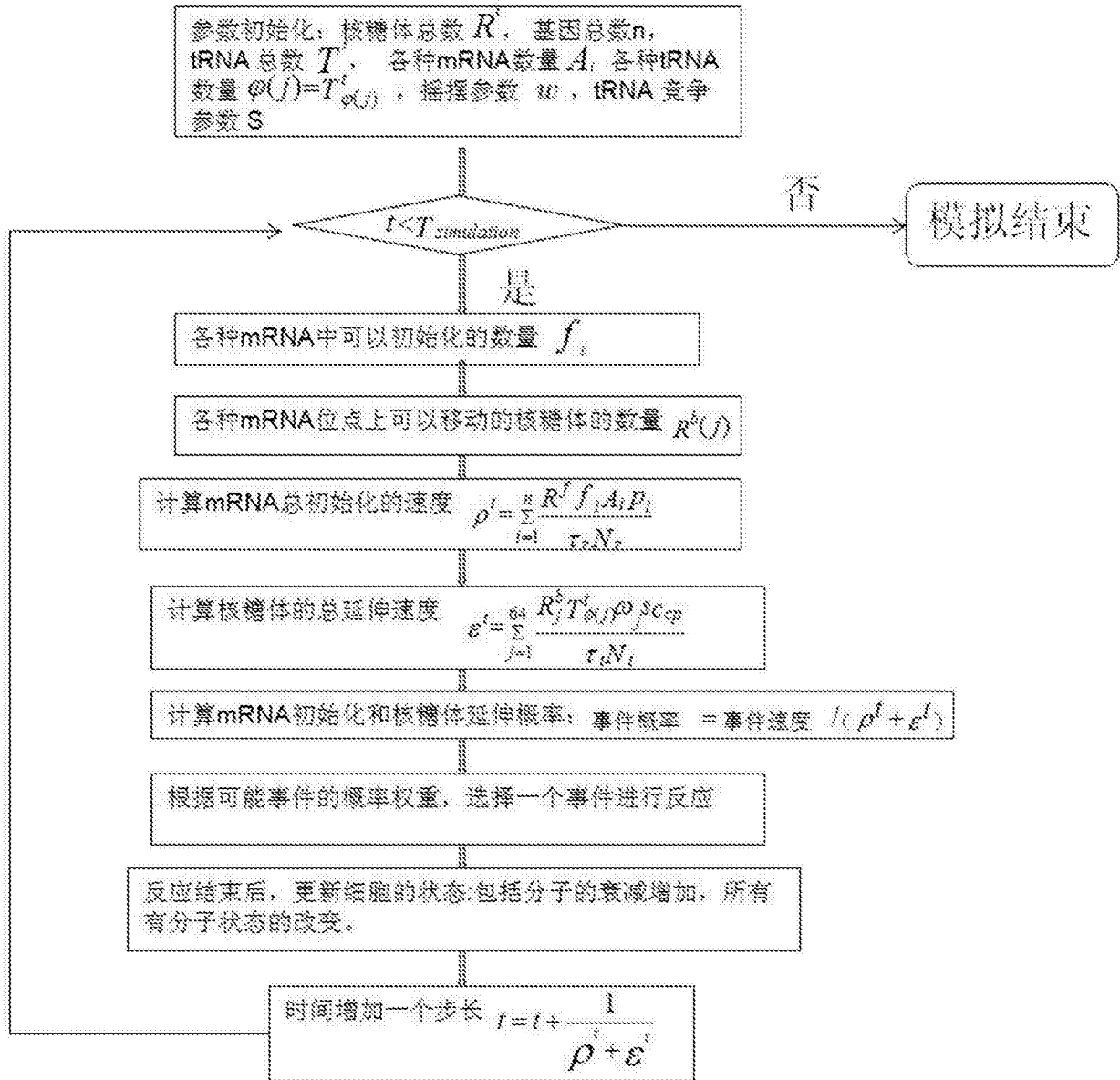


图1

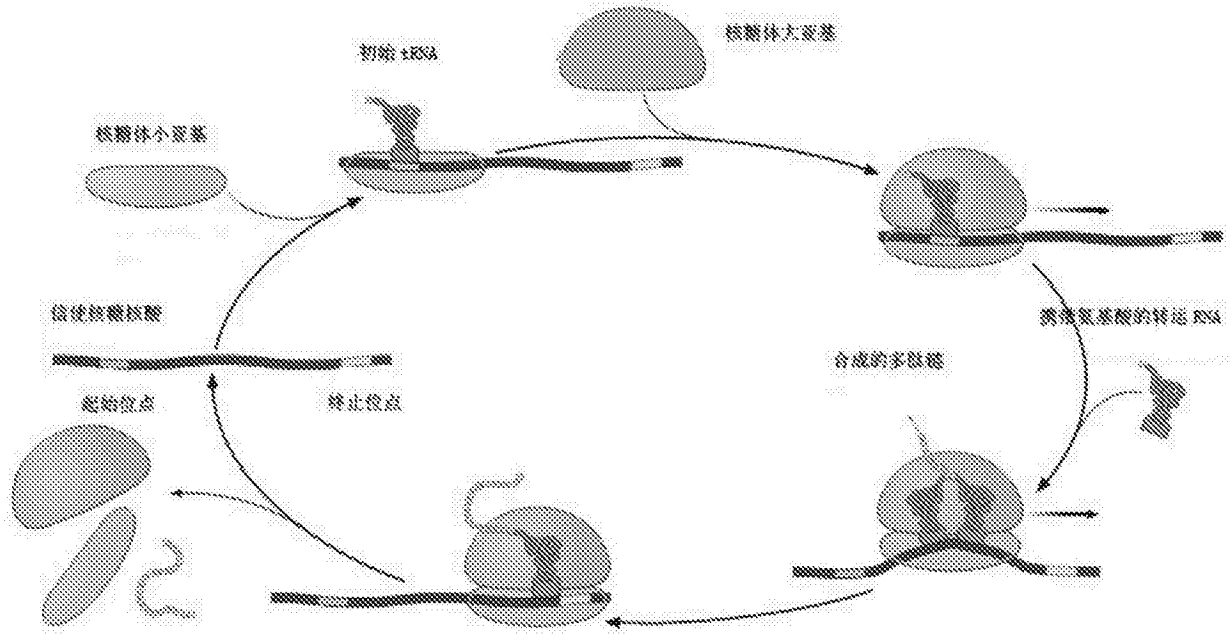


图2

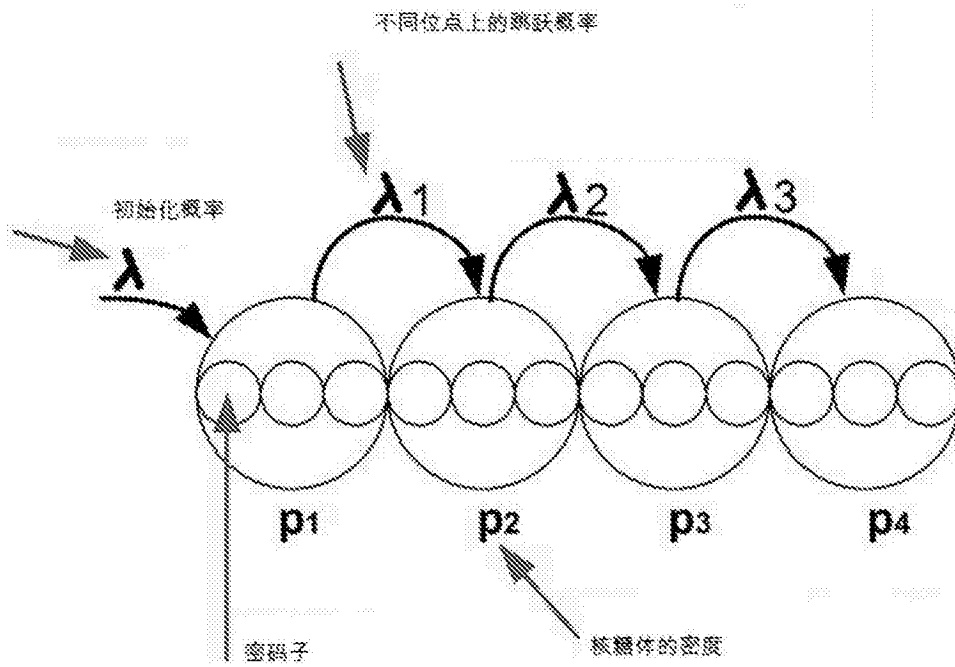


图3

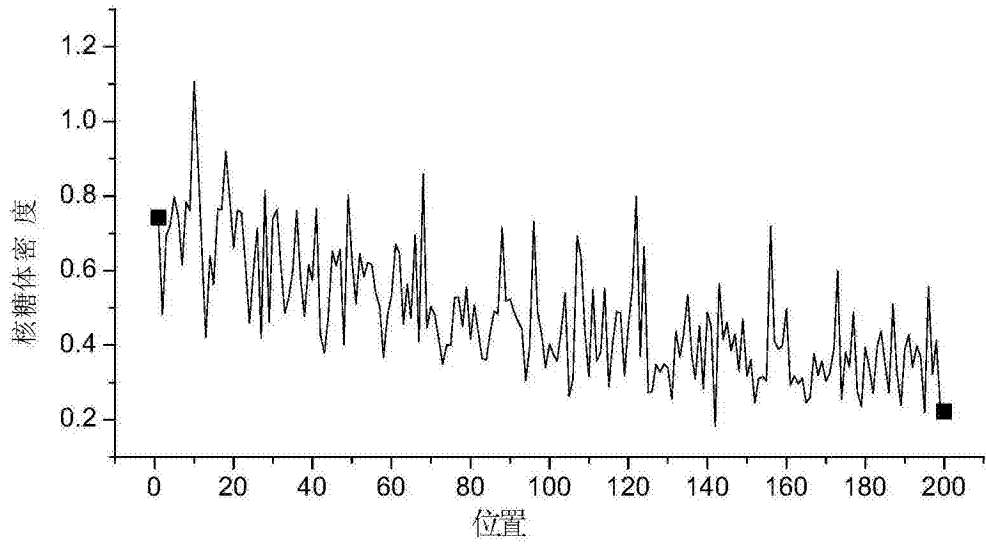


图4

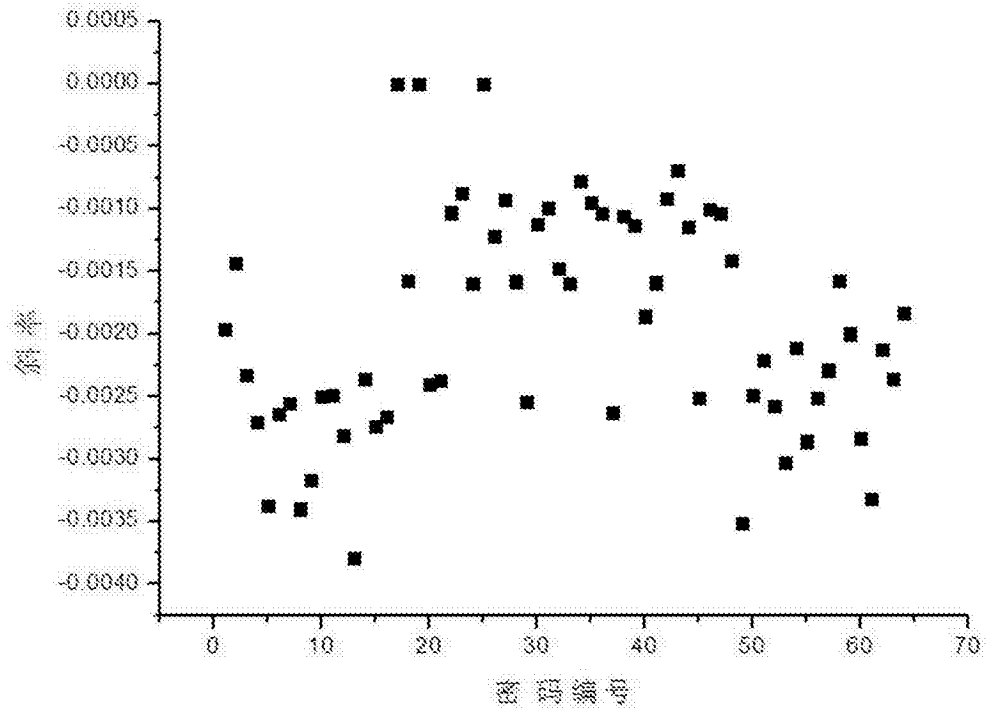


图5



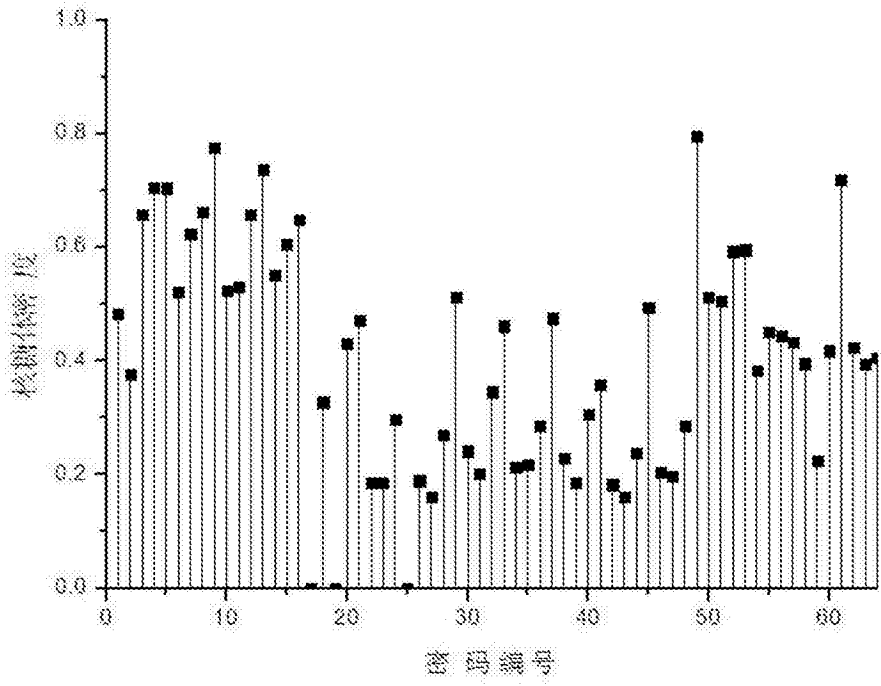


图6



图7