



## (12)发明专利

(10)授权公告号 CN 103106352 B

(45)授权公告日 2017.04.26

(21)申请号 201310034671.5

(22)申请日 2013.01.22

(65)同一申请的已公布的文献号  
申请公布号 CN 103106352 A

(43)申请公布日 2013.05.15

(73)专利权人 武汉科技大学  
地址 430080 湖北省武汉市青山区和平大道947号

(72)发明人 张凯 许进 胡威 石晓龙  
潘林强 张晓龙 符海东 强小利  
石心竹 贾峤

(74)专利代理机构 北京市柳沈律师事务所  
11105

代理人 沙捷

(51)Int.Cl.

G06F 19/20(2011.01)

(56)对比文件

CN 102063643 A, 2011.05.18,

KR 20100031218 A, 2010.03.22,

王延峰.DNA计算中的编码理论与方法研究.  
《中国博士学位论文全文数据库(基础科学辑)》  
.2009,第8,44-47页.

王延峰.DNA计算中的编码理论与方法研究.  
《中国博士学位论文全文数据库(基础科学辑)》  
.2009,第8,44-47页.

张凯.DNA编码问题及其复杂性研究.《计算机应用研究》.2008,第25卷(第11期),第3264-3267页.

审查员 贾云杰

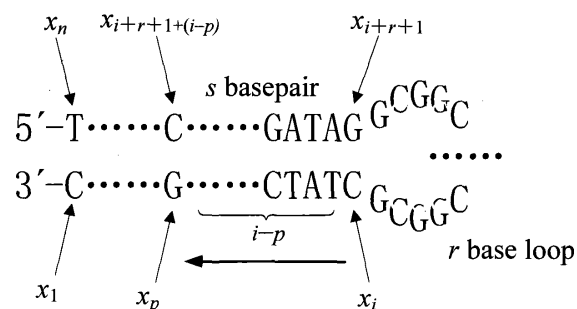
权利要求书2页 说明书5页 附图1页

(54)发明名称

一种全局最优化DNA计算序列编码方法

(57)摘要

本发明提供一种全局最优化DNA计算序列编码方法,包括:步骤1,利用剪枝策略搜索 $4^n$ 个DNA分子解空间;步骤2,将一个或者多个DNA计算的生物约束转换为对应的数学约束;步骤3,将该数学约束应用到该DNA分子解空间,产生适用于DNA计算的单链DNA序列集合。



1. 一种全局最优化DNA计算序列编码方法,包括:

步骤1,利用剪枝策略搜索 $4^n$ 个DNA分子解空间;

步骤2,将多个DNA计算的生物约束转换为对应的数学约束;

步骤3,将该数学约束应用到该DNA分子解空间,产生适用于DNA计算的单链DNA序列集合,

其中,步骤2中,所述生物约束至少包括发卡结构约束、连续性约束和GC含量约束;其中,所述发卡结构约束表示为:给定发卡环部的碱基数目 $r$ ,发卡茎部互补碱基数为 $s$ ,发卡茎部的碱基长度为 $i-p$ ;要求DNA序列 $X$ 反向形成环状结构后,茎部互补的碱基小于 $s$ ,

$$Q_{\text{发卡结构}} = \min\{p \mid \sum_{i=1}^p wc(x_i, x_{i+r+1}) \geq s, 1 \leq p \leq n, wc(x_i, x_{i+r+1}) = \begin{cases} 1, x_i = \overline{x_{i+r+1}} \\ 0, \text{else} \end{cases} \};$$

所述连续性约束表示为:给定连续性目标值 $Con_{\max}$ ,则要求DNA序列集合中任意单链DNA序列 $X$ 的连续性都小于该目标值,

$$Q_{\text{连续性}} = \min\{p \mid Con(x_i) \sum_{i=1}^p Con(x_i) \geq Con_{\max} - 1, 1 \leq p \leq n, Con(x_i) = \begin{cases} 1, x_i = x_{i-1} \\ 0, \text{else} \end{cases} \};$$

所述GC含量约束表示为:给定GC含量的目标范围 $[GC_{\min}, GC_{\max}]$ ,则 $GC_{\min} \leq \sum_{i=1}^k GC(x_i) \leq GC_{\max}$ 与

其等价的AT含量关系为 $AT_{\min} \leq \sum_{i=1}^k AT(x_i) \leq AT_{\max}$ ,其中 $AT_{\max} = n - GC_{\min}$ ,  $AT_{\min} = n - GC_{\max}$ :

$$Q_{\text{GC含量}} = \min\{p \mid \sum_{i=1}^p GC(x_i) > GC_{\max}, 1 \leq p \leq n, GC(x_i) = \begin{cases} 1, x_i = G \text{ or } C \\ 0, \text{else} \end{cases} \} \cup \\ \min\{p \mid \sum_{i=1}^p AT(x_i) > n - GC_{\min}, 1 \leq p \leq n, AT(x_i) = \begin{cases} 1, x_i = A \text{ or } T \\ 0, \text{else} \end{cases} \};$$

所述步骤2中,所述生物约束还包括:汉明距离约束、相似度约束、H-measure约束、反补汉明距离约束、自补汉明距离约束、解链温度约束、化学自由能约束、限定子序列约束和/或DNA模板序列约束。

2. 根据权利要求1所述的全局最优化DNA计算序列编码方法,其中,步骤2中,该数学约束为整数线性规划的条件不等式。

3. 根据权利要求1所述的全局最优化DNA计算序列编码方法,其中,步骤3中,基于该数学约束评价DNA分子解空间中的DNA序列是否满足编码约束。

4. 根据权利要求1所述的方法,其中,汉明距离约束表示为:

对于每一条DNA分子序列,给定汉明距离目标值 $H_{\text{target}}$ ,要求DNA序列集合中任意两条序列 $X$ 和 $Y$ 的汉明距离都大于该目标值;

$$Q_{\text{汉明距离}} = \{p \mid \sum_{i=1}^p h(x_i, y_i) \geq H_{\text{target}}, 1 \leq i, p \leq n, h(x_i, y_i) = \begin{cases} 0, x_i = y_i \\ 1, \text{else} \end{cases} \};$$

其中,相似度约束表示为:对于每一条DNA分子序列,相似度约束值 $\text{Similarity}_{\text{target}}$ ,则DNA序列 $X$ 和 $Y$ 之间,对于距离为 $k$ 的每次滑动,其汉明距离必须大于等于该约束:

$$Q_{\text{相似度}} = \{p \mid \sum_{i=1}^p h(x_i, \sigma^k(y_i)) \geq \text{Similarity}_{\text{target}}, 1 \leq i, p \leq n, h(x_i, y_i) = \begin{cases} 0, x_i = y_i \\ 1, \text{else} \end{cases} \}.$$

5. 根据权利要求1所述的全局最优化DNA计算序列编码方法, 其中, H-measure约束表示为: 给定H-measure约束值H-measure<sub>target</sub>, 则DNA序列X和Y之间, 对于距离为k的每次滑动, 其H-measure必须大于等于该约束值,

$$Q_{\text{H-measure}} = \{p \mid \sum_{i=1}^p bp(x_i, \sigma^k(y_i)) \geq H\text{-measure}_{\text{target}}, 1 \leq i, p \leq n, bp(x_i, y_i) = \begin{cases} 0, x_i = \bar{y}_i \\ 1, \text{else} \end{cases} \};$$

其中, 反补汉明距离约束表示为: 给定反补汉明距离的目标值RCHamming<sub>target</sub>, 则DNA序列集合中任意序列X和Y<sup>RC</sup>的汉明距离都大于该目标值, 令W=Y<sup>RC</sup>, W=5'-w<sub>1</sub>w<sub>2</sub>...w<sub>n-3</sub>', 则X和W的汉明距离大于该目标值,

$$Q_{\text{反补汉明距离}} = \{p \mid \sum_{i=1}^p h(x_i, w_i) \geq RCHamming_{\text{target}}, 1 \leq i, p \leq n, h(x_i, w_i) = \begin{cases} 0, x_i = w_i \\ 1, \text{else} \end{cases} \}.$$

6. 根据权利要求1所述的全局最优化DNA计算序列编码方法, 其中, 自补汉明距离约束表示为: DNA序列X和X<sup>R</sup>之间, 对于距离为k的每次滑动, 自补汉明距离目标值SCHamming<sub>target</sub>必须大于等于该约束, 令W=X<sup>R</sup>, W=3'-x<sub>n</sub>...x<sub>2</sub>x<sub>1</sub>-5', 则X和W的自补汉明距离大于该目标值

$$Q_{\text{自补汉明距离}} = \{p \mid \sum_{i=1}^p bp(x_i, \sigma^k(w_i)) \geq SCHamming_{\text{target}}, 1 \leq i, p \leq n, bp(x_i, w_i) = \begin{cases} 0, x_i = \bar{w}_i \\ 1, \text{else} \end{cases} \}.$$

7. 根据权利要求1所述的全局最优化DNA计算序列编码方法, 其中, 解链温度约束表示候选序列X和其补链X<sup>C</sup>的解链温度是否在给定上下限范围内, 给定解链温度T<sub>m</sub>的目标范围T<sub>mmin</sub><T<sub>m</sub><T<sub>mmax</sub>,

$$Q_{\text{解链温度}} = \min \{p \mid 4 \sum_{i=1}^p GC(x_i) + 2 \sum_{i=1}^p AT(x_i) \leq T_{m\max}, 1 \leq p \leq n, GC(x_i) = \begin{cases} 1, x_i = G \text{ or } C \\ 0, \text{else} \end{cases} \} \cup \\ \min \{p \mid 4 \sum_{i=1}^p GC(x_i) + 2 \sum_{i=1}^p AT(x_i) + 4(n-p) \geq T_{m\min}, AT(x_i) = \begin{cases} 1, x_i = A \text{ or } T \\ 0, \text{else} \end{cases} \}.$$

其中, 化学自由能约束表示为: 对于目标自由能ΔG<sub>min</sub>, 候选DNA序列和所有已选DNA序列的杂交自由能必须大于该自由能变化, 即ΔG>ΔG<sub>min</sub>

$$Q_{\text{最小自由能}} = \min \{p \mid \sum_{i=1}^p \Delta G \leq \Delta G_{\min}, 1 \leq i, p \leq n, \}.$$

8. 根据权利要求1所述的全局最优化DNA计算序列编码方法, 其中, 限定子序列约束表示为给定一组禁止出现的子序列, DNA序列可行解不包含这些核酸子序列片段, 假设Y=5'-y<sub>1</sub>y<sub>2</sub>...y<sub>m</sub>-3'代表这些核酸子序列, Length<sub>SubLength</sub>为其长度,

$$Q_{\text{禁止子序列}} = \min \{p \mid s(x_i, \sigma^k(y_i)) \sum_{i=1}^p s(x_i, \sigma^k(y_i)) \geq \text{Length}_{\text{subsequence}}, s(x_i, y_i) = \begin{cases} 1, x_i = y_i \\ 0, \text{else} \end{cases} \};$$

其中, DNA模板序列约束表示为: 给定模版序列, 该模版序列在固定的位置具有特定的碱基, DNA序列可行解符合该模版序列,

$$Q_{\text{模板序列}} = \min \{p \mid x_p \neq y_p, 1 \leq p \leq n, y_p \notin \{\text{free variables}\} \}.$$

## 一种全局最优化DNA计算序列编码方法

### 技术领域

[0001] 本发明涉及计算编码技术,更具体地,本发明涉及一种全局最优化DNA计算序列编码方法。

### 背景技术

[0002] 在DNA计算的过程中,单链DNA分子在溶液中任意扩散,溶液中会同时存在单链DNA分子及其三种相关的DNA分子:DNA反链、DNA补链、DNA反补链。它们都可能参与DNA计算的杂交反应,其中,X表示单链DNA序列, $X^C$ 表示X的Watson-Crick补序列, $X^R$ 表示X的反向序列, $X^{RC}$ 表示X的反补序列,如图1所示。

[0003] 杂交反应是DNA计算中最主要、最核心的反应,杂交反应具有方向性。通常,5'→3'的单链DNA分子和3'→5'的单链DNA分子在氢键的作用下相互绑定形成双螺旋结构。

[0004] 杂交分为特异性杂交和非特异性杂交,特异性杂交为单链DNA分子X和其补链 $X^C$ 的碱基对完全互补的杂交反应,除此之外的其它杂交反应均为非特异性杂交。需要强调的是,不能认为单链DNA分子X和其补链 $X^C$ 必定发生特异性杂交,它们通常发生特异性杂交,但如果滑动后有足够多的碱基互补时,它们之间也可能会发生非特异性杂交,如图2所示。

[0005] 目前,DNA计算的基本问题是提高DNA计算的可靠性、有效性和可扩充性。随着DNA计算求解问题规模的不断增大,需要使用的DNA序列数量急剧增多;为了保证DNA计算的有效性和可靠性,需要严格的编码约束以提高编码质量。DNA序列越长,其合成的成本就越高,且在试验中控制DNA长链的难度会越大。这样,DNA编码问题的目标为:DNA序列长度尽可能短;编码约束尽可能强;编码数量尽可能多。

[0006] 然而,这些目标相互存在着冲突。对于给定长度的DNA编码,编码约束越强则编码数量就越少;DNA链长越短,满足条件的DNA数量也越少。这种多目标性使得DNA编码问题的复杂性和计算量急剧增加。

[0007] DNA编码受到碱基距离约束、化学热力学约束、DNA二级结构约束、DNA分子组成约束等多种编码规则的组约束。在计算量上是NP完全问题,即随着问题规模的增大,对于求解最优DNA序列集合的计算量呈指数增长。如果设计链长为n的一组DNA序列,由于每位的碱基有{A,T,G,C}四种可能,其解空间为 $4^n$ 。如果要求DNA序列集合最大,仅考虑DNA序列之间非特异性杂交的限制,对于长度为n的DNA序列,就等价于求解 $4^n$ 个顶点的图的最大独立集问题。如果还考虑DNA二级结构等其它约束,问题就变得更加复杂。

[0008] 当前各种DNA序列设计算法适用的编码约束有限,例如Feldkamp提出的算法适用4种编码约束,Frutos提出的模版映射算法仅适用7种编码约束,Shin等提出的多目标进化算法最大也仅能适用于9种编码约束。

[0009] 因此,这些算法对不同的DNA计算模型不具备通用性,此外这些非精确算法难以高效的设计出满足约束的最大DNA序列集合,无法求解大规模的DNA计算问题实例。

### 发明内容

[0010] 为克服现有技术的上述缺陷,本发明提出一种全局最优化DNA计算序列编码方法。

[0011] 根据本发明的一个方面,提出了一种全局最优化DNA计算序列编码方法,包括:步骤1,利用剪枝策略搜索 $4^n$ 个DNA分子解空间;步骤2,将一个或者多个DNA计算的生物约束转换为对应的数学约束;步骤3,将该数学约束应用到该DNA分子解空间,产生适用于DNA计算的单链DNA序列集合。

[0012] 其中,步骤2中,该数学约束为整数线性规划的条件不等式。

[0013] 步骤2中,该一个或者多个生物约束包括:汉明距离约束、相似度约束、H-measure约束、反补汉明距离约束、自补汉明距离约束、发卡结构约束、连续性约束、GC含量约束、解链温度约束、化学自由能约束、限定子序列约束和/或DNA模板序列约束。

[0014] 其中,步骤3中,基于该数学约束评价DNA分子解空间中的DNA序列是否满足编码约束。

[0015] 本发明将DNA计算的生物约束转换为数学约束,将12种DNA编码约束转换为整数线性规划的条件不等式,产生并评价DNA序列是否满足编码约束。采用上述技术方案,可高效地搜索 $4^n$ 的DNA分子解空间,生成满足约束且更大数量规模的DNA分子集合;在保证DNA计算过程的有效性和可靠性的前提下,提高可求解问题的规模。

## 附图说明

[0016] 图1示出单链DNA分子相关序列示意图;

[0017] 图2示出DNA特异性杂交和非特异性杂交的示意图;

[0018] 图3示出发卡结构示意图。

[0019] 如图所示,为了能明确实现本发明的实施例的结构,在图中标注了特定的结构和运行方式,但这仅为示意需要,并非意图将本发明限定在该特定方式、器件和环境中,根据具体需要,本领域的普通技术人员可以将这些器件和环境进行调整或者修改,所进行的调整或者修改仍然包括在后附的权利要求的范围内。

## 具体实施方式

[0020] 下面结合附图和具体实施例对本发明提供的一种全局最优化DNA计算序列编码方法进行详细描述。

[0021] 在以下的描述中,将描述本发明的多个不同的方面,然而,对于本领域内的普通技术人员而言,可以仅仅利用本发明的一些或者全部结构或者流程来实施本发明。为了解释的明确性而言,阐述了特定的数目、配置和顺序,但是很明显,在没有这些特定细节的情况下也可以实施本发明。在其他情况下,为了不混淆本发明,对于一些众所周知的特征将不再进行详细阐述。

[0022] 本发明公开了一种全局最优化DNA计算序列编码方法,适用于DNA计算的单链DNA序列集合。该方法利用剪枝策略高效地搜索 $4^n$ 的DNA序列解空间,生成满足给定约束条件的最大的DNA序列集合。该方法将12种DNA计算的生物约束,转换为整数线性规划的条件不等式,生成DNA序列集合。该方法设计的DNA分子在DNA计算的过程中不易发生错配,能提高DNA计算的有效性,可靠性。该方法基于全局最优化,能比局部最优化方法设计出更多的DNA序列,以求解更大规模DNA计算问题实例。

[0023] 具体地,该一种全局最优化DNA计算序列编码方法,包括:利用剪枝策略搜索 $4^n$ 个DNA分子解空间;将一个或者多个DNA计算的生物约束转换为对应的数学约束;将该数学约束应用到该DNA分子解空间,产生适用于DNA计算的单链DNA序列集合;该数学约束为整数线性规划的条件不等式。

[0024] 其中,12种DNA计算的生物约束包括汉明距离约束、相似度约束、H-measure约束、反补汉明距离约束、自补汉明距离约束、发卡结构约束、连续性约束、GC含量约束、解链温度约束、化学自由能约束、限定子序列约束和DNA模板序列约束。

[0025] 该12种生物约束对应的数学约束表示式分别如下:

[0026] 汉明距离约束

[0027] 本实施例提供的每一条DNA分子序列,给定汉明距离目标值 $H_{\text{target}}$ ,则要求DNA序列集合中任意两条序列X和Y的汉明距离都大于该目标值。

$$[0028] \quad Q_{\text{汉明距离}} = \{p | \sum_{i=1}^p h(x_i, y_i) \geq H_{\text{target}}, 1 \leq i, p \leq n, h(x_i, y_i) = \begin{cases} 0, x_i = y_i \\ 1, \text{else} \end{cases} \} \quad (1)$$

[0029] 相似度约束

[0030] 本实施例提供的每一条DNA分子序列,给定相似度约束值 $\text{Similarity}_{\text{target}}$ ,则DNA序列X和Y之间,对于距离为k的每次滑动,其汉明距离必须大于等于该约束。

$$[0031] \quad Q_{\text{相似度}} = \{p | \sum_{i=1}^p h(x_i, \sigma^k(y_i)) \geq \text{Similarity}_{\text{target}}, 1 \leq i, p \leq n, h(x_i, y_i) = \begin{cases} 0, x_i = y_i \\ 1, \text{else} \end{cases} \} \quad (2)$$

[0032] H-measure约束

[0033] 给定H-measure约束值 $H\text{-measure}_{\text{target}}$ ,则DNA序列X和Y之间,对于距离为k的每次滑动,其H-measure必须大于等于该约束值。

$$[0034] \quad Q_{\text{H-measure}} = \{p | \sum_{i=1}^p bp(x_i, \sigma^k(y_i)) \geq H\text{-measure}_{\text{target}}, 1 \leq i, p \leq n, bp(x_i, y_i) = \begin{cases} 0, x_i = \overline{y_i} \\ 1, \text{else} \end{cases} \} \quad (3)$$

[0035] 反补汉明距离约束

[0036] 给定反补汉明距离的目标值 $RCHamming_{\text{target}}$ ,则要求DNA序列集合中任意序列X和 $Y^{\text{RC}}$ 的汉明距离都大于该目标值。令 $W=Y^{\text{RC}}, W=5'-w_1w_2\cdots w_{n-3}'$ ,则X和W的汉明距离大于该目标值。

$$[0037] \quad Q_{\text{反补汉明距离}} = \{p | \sum_{i=1}^p h(x_i, w_i) \geq RCHamming_{\text{target}}, 1 \leq i, p \leq n, h(x_i, w_i) = \begin{cases} 0, x_i = w_i \\ 1, \text{else} \end{cases} \} \quad (4)$$

[0038] 自补汉明距离约束

[0039] 给定自补汉明距离目标值 $SCHamming_{\text{target}}$ ,则DNA序列X和 $X^{\text{R}}$ 之间,对于距离为k的每次滑动,自补汉明距离必须大于等于该约束。令 $W=X^{\text{R}}, W=3'-x_n\cdots x_2x_1-5'$ ,则X和W的自补汉明距离大于该目标值。

$$[0040] \quad Q_{\text{自补汉明距离}} = \{p | \sum_{i=1}^p bp(x_i, \sigma^k(w_i)) \geq SCHamming_{\text{target}}, 1 \leq i, p \leq n, bp(x_i, w_i) = \begin{cases} 0, x_i = \overline{w_i} \\ 1, \text{else} \end{cases} \} \quad (5)$$

[0041] 发卡结构约束

[0042] 单链DNA分子产生二级结构通常由自身的反向折叠形成,发卡结构为典型的自身折叠结构,如图3所示。

[0043] 给定发卡环部的碱基数目 $r$ ,发卡茎部互补碱基数为 $s$ ,发卡茎部的碱基长度为 $i-p$ 。要求DNA序列 $X$ 反向形成环状结构后,茎部互补的碱基小于 $s$ 。

$$[0044] \quad Q_{\text{发卡结构}} = \min\{p \mid \sum_{i=1}^p wc(x_i, x_{i+r+1}) \geq s, 1 \leq p \leq n, wc(x_i, x_{i+r+1}) = \begin{cases} 1, x_i = \overline{x_{i+r+1}} \\ 0, \text{else} \end{cases} \} \quad (6)$$

[0045] 连续性约束

[0046] DNA序列某碱基连续出现会导致不期望的二级结构,在分子生物实验中通常需要避免连续碱基序列“AAA”,“CCC”,“GGG”,“TTT”。给定连续性目标值 $Con_{\max}$ ,则要求DNA序列集合中任意单链DNA序列 $X$ 的连续性都小于该目标值。

$$[0047] \quad Q_{\text{连续性}} = \min\{p \mid Con(x_i) \sum_{i=1}^p Con(x_i) \geq Con_{\max} - 1, 1 \leq p \leq n, Con(x_i) = \begin{cases} 1, x_i = x_{i-1} \\ 0, \text{else} \end{cases} \} \quad (7)$$

[0048] GC含量约束

[0049] 给定GC含量的目标范围 $[GC_{\min}, GC_{\max}]$ ,则 $GC_{\min} \leq \sum_{i=1}^k GC(x_i) \leq GC_{\max}$ ,与其等价的AT含量关系为 $AT_{\min} \leq \sum_{i=1}^k AT(x_i) \leq AT_{\max}$ ,其中 $AT_{\max} = n - GC_{\min}$ ,  $AT_{\min} = n - GC_{\max}$ 。

$$[0050] \quad Q_{\text{GC含量}} = \min\{p \mid \sum_{i=1}^p GC(x_i) > GC_{\max}, 1 \leq p \leq n, GC(x_i) = \begin{cases} 0, x_i = G \text{ or } C \\ 1, \text{else} \end{cases} \} \cup (8)$$

$$[0051] \quad \min\{p \mid \sum_{i=1}^p AT(x_i) > n - GC_{\min}, 1 \leq p \leq n, AT(x_i) = \begin{cases} 0, x_i = A \text{ or } T \\ 1, \text{else} \end{cases} \}$$

[0052] 解链温度约束

[0053] 解链温度约束用来检查候选序列 $X$ 和其补链 $X^c$ 的解链温度是否在给定上下限范围内,给定解链温度 $T_m$ 的目标范围 $T_{m\min} < T_m < T_{m\max}$ 。

$$[0054] \quad Q_{\text{解链温度}} = \min\{p \mid 4 \sum_{i=1}^p GC(x_i) + 2 \sum_{i=1}^p AT(x_i) \leq T_{m\max}, 1 \leq p \leq n, GC(x_i) = \begin{cases} 0, x_i = G \text{ or } C \\ 1, \text{else} \end{cases} \} \cup (9)$$

$$[0055] \quad \min\{p \mid 4 \sum_{i=1}^p GC(x_i) + 2 \sum_{i=1}^p AT(x_i) + 4(n-p) \geq T_{m\min}, AT(x_i) = \begin{cases} 0, x_i = A \text{ or } T \\ 1, \text{else} \end{cases} \}$$

[0056] 化学自由能约束

[0057] 给定目标自由能 $\Delta G_{\min}$ ,候选DNA序列和所有已选DNA序列的杂交自由能必须大于该自由能变化,即 $\Delta G > \Delta G_{\min}$ 。

$$[0058] \quad Q_{\text{最小自由能}} = \min\{p \mid \sum_{i=1}^p \Delta G \leq \Delta G_{\min}, 1 \leq i, p \leq n, \} \quad (10)$$

[0059] 限定子序列约束

[0060] 给定一组禁止出现的子序列,例如限制性核酸内切酶EcoRI (5'-GAATTC-3') 和HindIII (5'-AAGCTT-3'),DNA序列可行解都必须不包含这些核酸子序列片段。假设 $Y = 5'-y_1 y_2 \cdots y_m - 3'$ 代表这些核酸子序列,LengthSubLength为其长度。

$$[0061] \quad Q_{\text{禁止子序列}} = \min\{p \mid s(x_i, \sigma^k(y_i)) \sum_{i=1}^p s(x_i, \sigma^k(y_i)) \geq \text{Length}_{\text{subsequence}}, s(x_i, y_i) = \begin{cases} 1, x_i = y_{i-1} \\ 0, \text{else} \end{cases} \} \quad (11)$$

[0062] DNA模板序列约束

[0063] 给定一个模版序列,则该模版序列在固定的位置具有特定的碱基,因此DNA序列可

行解都必须符合该模版序列。

$$[0064] \quad Q_{\text{模板序列}} = \min\{p \mid x_p \neq y_p, 1 \leq p \leq n, y_p \notin \{\text{free variables}\}\} \quad (12)$$

[0065] 以上12种适应度函数可以根据需要,任意选择若干个,候选序列若依次通过这些约束条件即为可行解,可以用于DNA计算的试验。

[0066] 最后应说明的是,以上实施例仅用以描述本发明的技术方案而不是对本技术方法进行限制,本发明在应用上可以延伸为其他的修改、变化、应用和实施例,并且因此认为所有这样的修改、变化、应用、实施例都在本发明的精神和教导范围内。



