

基于自组装模型的最大团问题 DNA 计算算法

李肯立 罗兴 吴帆 周旭 黄鑫

(湖南大学信息科学与工程学院 长沙 410082)

(lk1510@263.net)

An Algorithm in Tile Assembly Model for Maximum Clique Problem

Li Kenli, Luo Xing, Wu Fan, Zhou Xu, and Huang Xin

(School of Information Science and Engineering, Hunan University, Changsha 410082)

Abstract DNA computing is an efficient method to solve computational problems, especially for NP complete problems that cannot be efficiently solved using the traditional Turing machine. However, with the development of DNA computation, it presently suffers from several challenging problems such as relatively high error rates, DNA space explosion, etc. How to decrease error rates has become an important issue for DNA computation. To solve the maximum clique problem which is an NP complete problem, with low DNA hybridization error rates, this paper introduces a DNA self-assembly model and presents a novel DNA computing algorithm. This algorithm decreases error rates by decreasing experiment operations. Moreover, the DNA structures of initial molecular, regular molecular and detection molecular are designed, and encoding scheme of DNA molecular is described, too. The proposed algorithm needs $\Theta(n + |E|)$ types of tiles and $\Theta(1)$ times of experiment operations, where n and $|E|$ are the numbers of vertex and edge of a graph respectively. According to the analysis, our algorithm can not only improve the accuracy of the solutions, but also greatly reduce the complexity of the experiment, which leads to easier experiment operability, as compared with the other DNA models including sticker model, insert-delete system, etc.

Key words DNA computing; self-assembly; parallel computing; NP-complete problem; maximum clique problem

摘要 DNA 计算在解决 NP 完全问题时,有着传统图灵机无法比拟的优势.但是随着 DNA 计算研究的不断深入,传统 DNA 计算模型显现出杂交错率率和生化操作复杂性过高的缺点.如何提高 DNA 计算结果的准确性在 DNA 计算研究中日益重要.针对 NP 完全的最大团问题,引入 DNA 自组装模型,提出了一种求解最大团问题的 DNA 计算算法.算法通过减少实验的操作步骤数,以降低生化解的错误率,给出了 DNA 分子的编码方案及结果检测的实验方法.算法设计的 tiles 种类为 $\Theta(n+|E|)$,生化操作复杂性为 $\Theta(1)$,其中 n 为图的顶点数, $|E|$ 为边数.与求解最大团问题的其他 DNA 算法的对比分析表明,本算法不仅明显提高了生化解的准确性,且算法的生化实验复杂度低,具有良好的实验操作性.

关键词 DNA 计算;自组装;并行计算;NP 完全问题;最大团问题

中图法分类号 TP301.6

收稿日期:2011-08-14;修回日期:2012-04-09

基金项目:国家自然科学基金项目(61133005,61070057,61202109,61173013);国家科技支撑计划基金项目(2012BAH09B02);湖南省杰出青年基金项目(12JJ011)

1994年,Adleman首次运用DNA分子解决7节点的Hamilton路径问题^[1],提供了一种以生物技术来实现电子计算机在多项式时间内难以解决的困难问题的计算方式,并且证明了运用DNA分子进行计算的可行性.1996年,Roweis等人提出经典的粘贴DNA计算模型^[2];1996年,Kari和Thierrin引入了建立在上下文删除基础上的插入-删除系统^[3];1998年,Winfree等人构建了自组装模型^[4];1999年,由Jonoska等人提出 k -臂DNA计算模型^[5];2000年,Head等人提出质粒DNA计算模型等^[6].10多年来,DNA计算无论是在理论模型上,还是实验与检测技术上都取得了惊人的进步.

1963年,Wang从数学角度以tiles作为计算指令,指出tiles可以通过自组装过程在二维平面形成周期性格局,且证明了该过程具有图灵等价性,为DNA自组装模型提供了理论基础^[7].20世纪90年代,Seeman以DNA分子作为Wang的tiles,构建了许多较复杂的结构,比如四边形、环等^[8-9];1998年Winfree利用Wang的tiles形成二维晶体结构^[4],提出了以DNA tiles分子自组装的过程来实现计算的思想^[10].2008年,Yu等人应用DNA单链进一步组装形成了不同多面体结构^[11].

最大团问题是理论计算机科学中一个经典的NP完全性问题.先后有不少学者提出了最大团问题的DNA计算算法^[12-14].基于剪枝策略,我们提出过一种改进的最大团问题DNA计算算法^[15].该算法将求解 n 个顶点的最大团问题所需DNA分子链数从 $O(2^n)$ 减少至 $O(\sqrt{3}^n)$.但是算法的生化实验操作复杂度较高,很难保证实验结果的准确性.本文引入Winfree提出的DNA自组装模型,提出了一种解决最大团问题的新算法.算法不仅提高了生化解的准确性,且生化实验复杂度低,具有良好的实验操作性.

1 DNA自组装模型

1.1 自组装模型的二维平面结构

DNA链根据碱基互补配对规则自组装可形成一维、二维、三维结构.Adleman首次构建的DNA计算模型、粘贴模型、剪接模型等都是利用DNA链组装形成的一维结构.但是其生化实验步骤会随着问题规模的增加呈线性增加^[16],这就给生化实验带来了挑战.本文主要基于DNA链自组装形成的二

维平面结构,提出一种新的最大团问题DNA计算算法.

1998年,以Wang的tiles自组装理论为理论基础,Winfree设计了自组装单元DNA双交叉分子DX^[4].DX指在螺旋域之间包含2个交叉链接的DNA分子结构^[17].理论上可构建5种DX结构,但大量实验分析发现DAO(double-crossover antiparallel odd)和DAE(double-crossover antiparallel even)这两种结构具有稳定性^[4],分别如图1,2所示:

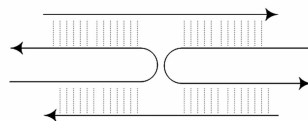


Fig. 1 Abstract structure of the DAO molecular.

图1 DAO分子抽象结构

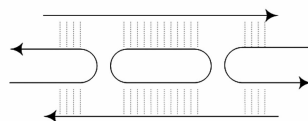


Fig. 2 Abstract structure of the DAE molecular.

图2 DAE分子抽象结构

观察图1,2可知,DAO结构包含4条DNA链,其中有2条链分布在2个双螺旋上,另外2条只是分布式在一个双螺旋上,在2个交叉点之间半螺旋为3个(奇数个).而DAE结构包含5条DNA链,其中有3条链分布在2个双螺旋上,另外2条链分布在1个双螺旋上,在2个交叉点之间半螺旋为4个(偶数个).

Wang的tiles自组装理论规定只能相同颜色的边才能结合形成较大的格局^[7].根据DNA序列的碱基互补配对的原则,DX只有当DX粘性末端的碱基序列互补时才能在相应的位置处结合形成较大的格局^[4].

1.2 自组装模型的数学描述

根据Wang的tiles自组装理论模型的形式化描述以及Brun引入DNA的tiles自组装模型^[18],结合最大团问题的特点,可将基于DAE结构的自组装模型表示为 $S = \langle T, g, \tau \rangle$,其中 T 为DAE分子集, g 为能量函数, τ 为常数.相应的符号描述如下:

1) Σ : DAE分子中所有粘性末端所表示的符号集,即所要参与计算的符号集.

2) $D' = \{UL, DL, UR, DR\}$:分别表示DAE分子的左上角、左下角、右上角、右下角4处的粘性

末端.

3) $\{\sigma_{UL}, \sigma_{DL}, \sigma_{UR}, \sigma_{DR}\} \in \Sigma^4$: DAE 分子的形式化描述, 表示 DAE 分子分别在 D' 中 4 个粘性末端表示的符号.

4) $(x, y) \in \mathbb{Z}^2$: 表示 DAE 块在二维平面上的位置.

5) $bd_d(t) \in \Sigma$, 其中 $t \in T, d \in D'$: 表示 DAE 分子 t 位置 d 所表示的符号.

6) 函数 $g: \Sigma \times \Sigma \rightarrow R$, 此函数的定义如下:

6.1) $\forall \sigma \in \Sigma$, 则 $g(\text{null}, \sigma) = 0$;

6.2) 若 σ 与 $\bar{\sigma}$ 互补, 则 $g(\sigma, \bar{\sigma}) = 1$.

7) 对于格局 C , 存在函数 $A_C(x, y): \mathbb{Z} \times \mathbb{Z} \rightarrow T$, 此函数有如下特点:

7.1) 若 $(x, y) \in \mathbb{Z}^2$, 且 $A_C(x, y) \neq \text{null}$, 则 $(x, y) \in C$;

7.2) 若只有有限个不同的 $(x, y) \in C$, 则 C 为一个有限集.

8) 若一个 DAE 块 t 能在格局 C 的 (x, y) 位置结合产生一个新的格局 C' , 需要满足以下条件:

8.1) $(x, y) \notin C$;

8.2) $\sum_{d \in D', d' \in D} g(bd_d(t), bd_{d'}(A_C(d'(x, y)))) \geq \tau$ (其中 d 与 d' 为相对的方向, 比如: UR 与 DL 相对);

8.3) $\forall (u, v) \in \mathbb{Z}^2, (u, v) \neq (x, y)$, 则 $A_{C'}(u, v) = A_C(u, v)$;

8.4) $A_C(x, y) = t$.

2 基于自组装模型的最大团问题 DNA 算法

2.1 最大团问题

设 G 是一个简单图, S 是 G 的一个非空的顶点子集. 如果由 S 在 G 中的导出子图 $G[S]$ 是图 G 的一个完全子图, 则称 S 是 G 的一个团. 设 S 是图 G 的一个团, 如果对于图 G 中的任意一个团 S' , 均有 $|S'| \leq |S|$, 则称 S 是 G 的一个最大团. 如何求解图的最大团? 这是一个著名的 NP 完全问题. 它不但在模式匹配、人工智能、计算机图像工程技术上有直接或者间接的应用, 而且在数学理论本身也具有好的应用. 因此, 给出快速准确的图的最大团算法具有重大意义.

多年来, 对最大团问题的求解算法进行了很多研究, 如常规算法、神经网络算法与遗传算法等^[19-21]. 但由于 NP 完全问题所固有的复杂性, 现有算法或

者仅适用于某些特殊图, 或具有指数时间的复杂度, 效率低, 缺乏实用性. 并且这些算法均是基于图灵机设计的, 因而不可能在克服问题的 NP 完全性上有实质突破. DNA 计算由于极大的存储密度和高度并行性等特点, 在解决 NP 难问题时有着独特的优势. 近年来, 不少学者提出了基于 DNA 计算模型的最大团问题算法. 虽然这些算法能够在多项式时间内求解, 但是具有较高的生化实验操作复杂度, 降低了生化解的准确性. 针对此问题, 本文引入 DNA 自组装模型, 根据 DX 的结构特性设计出相应的 DNA 分子. 结合基因工程的生化操作, 提出了基于自组装模型的最大团问题 DNA 算法. 该算法实验操作复杂度为常数级, 明显提高了生化解的准确性, 具有良好的实验操作性.

2.2 DX 分子设计方案

本文算法思想主要分为 3 个部分. 首先设计初始分子, 充分退火反应后生成顶点集所有子集的初始格局. 然后设计规则分子, 判断顶点子集是否任何两个顶点都邻接. 若顶点子集中顶点两两邻接, 则说明此顶点子集的导出子图为团. 最后设计检测分子, 通过凝胶电泳和荧光标记技术观察提取出最大团. 相应的生化操作描述如下.

1) 合并 (merge): 给定试管 P_1 和 P_2 , 操作 $\cup(P_1, P_2) = P_1 \cup P_2$ 表示将试管 P_1 和 P_2 合并到一个试管中而不改变 P_1 和 P_2 中的任何链.

2) 提取 (extract): 给定试管 P_1 和 P_2 , 操作 $P_2 = \text{extract}(P_1)$ 表示提取试管 P_1 中具有某种特征的 DNA 分子到试管 P_2 .

3) 退火 (anneal): 给定试管 P , 操作 $\text{anneal}(P)$ 将试管 P 中的所有 DNA 单链变成双链.

4) 荧光标记 (fluorescence-labeling): 给定试管 P , 该操作将具有某种特征的 DNA 分子作标记.

5) 凝胶电泳 (gel-electrophoresis): 给定试管 P , 该操作将试管中的 DNA 分子按照分子大小进行分离.

2.2.1 初始分子的设计

简单无向图 $G = \{V, E\}$, 其中 $V = \{v_1, v_2, \dots, v_n\}$ 为顶点集, E 为边集, n 为顶点个数. $(v_1^{X_1}, v_2^{X_2}, \dots, v_n^{X_n})$ 表示初始分子退火生成的初始格局, $X_i \in \{0, 1\}$, 当 $X_i = 1$ 时, 表示顶点子集包含顶点 v_i , 否则不包含 ($i \in \{1, 2, \dots, n\}$). 根据 DAE 结构的特点, 初始分子的设计如图 3 所示, 其中 s 和 e 是控制自组装反应的始端和终端, 与图 G 无关; $i \in \{1, 3, \dots, n-1\}$

$(3+n\%2)$ }, 其中 $n\%2$ 表示 n 对 2 的余数, 即若 n 为偶数时, $i \in \{1, 3, \dots, n-3\}$, n 为奇数时, $i \in \{1, 3, \dots, n-4\}$.

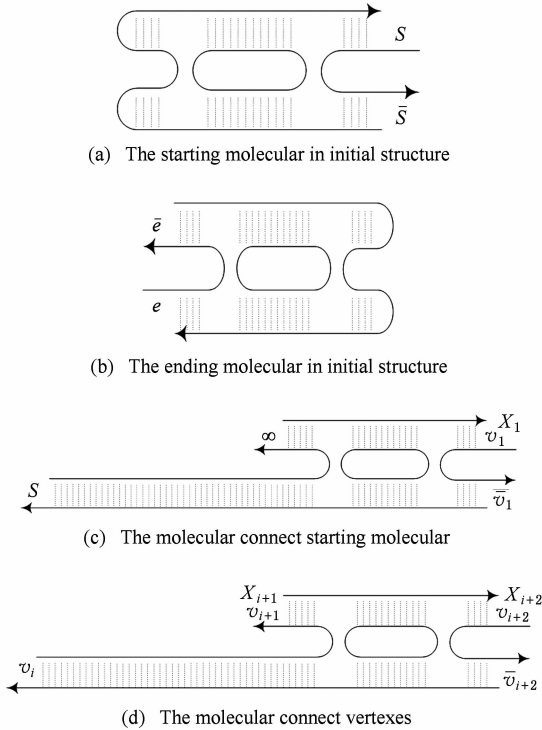


Fig. 3 Abstract structure of the initial DAE molecular.
图 3 初始分子 DAE 抽象结构

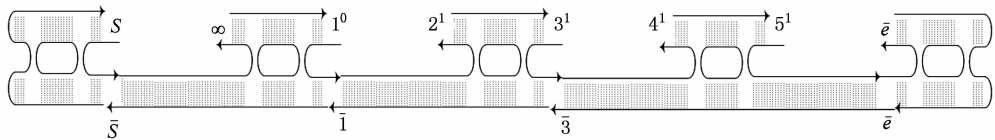


Fig. 6 The initial configuration of $\{v_1^0, v_2^0, v_3^0, v_4^0, v_5^0\}$ when $V = \{v_1, v_2, v_3, v_4, v_5\}$.
图 6 顶点集为 $\{v_1, v_2, v_3, v_4, v_5\}$ 时子集为 $\{v_1^0, v_2^0, v_3^0, v_4^0, v_5^0\}$ 的初始格局

2.2.2 规则分子的设计

规则分子的作用为判断初始分子退火生成的初始格局是否满足相应的条件. 对于最大团问题, 首先需要判断顶点子集的导出子图是否为团, 即判断该顶点子集的任意两顶点间是否邻接. 若顶点子集中的两个顶点邻接, 则存在如图 7 所示的 DAE 分

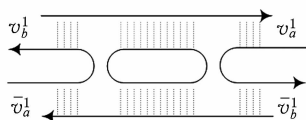


Fig. 7 Abstract structure of two adjacent vertices.
图 7 邻接顶点的 DAE 分子抽象结构

为了能准确生成所有顶点子集的初始状态结构, 分两种情况设计处理边界终端 e 的 DAE 分子, 分别如图 4, 5 所示. 故用于生成所有情况的初始格局的初始分子 DAE 种类为 $\Theta(n)$.

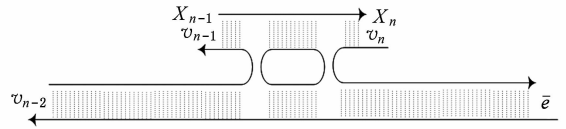


Fig. 4 Molecular form of e when n is odd.
图 4 n 为奇数时连接边界终端 e 的 DAE 分子

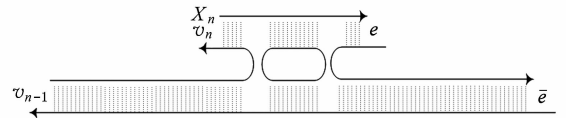


Fig. 5 Molecular form of e when n is even.
图 5 n 为偶数时连接边界终端 e 的 DAE 分子

根据顶点个数奇偶性, 按照上述规则设计相应的 DAE 分子, 运用 PCR 聚合酶链式反应对这 $\Theta(n)$ 类 DAE 分子迅速大量复制, 退火反应后生成所有的解空间. 如图 6 所示, 当顶点集为 $\{v_1, v_2, v_3, v_4, v_5\}$ 时, 子集为 $\{v_1^0, v_2^0, v_3^0, v_4^0, v_5^0\}$ 的初始分子退火生成的初始格局.

子, 综合边界情况把规则分子设计成如图 8 所示的形式.

图 8(a)(d) 的 DAE 分子控制所有顶点与始端 s 和终端 e 的反应, 图 8(b) 的 DAE 结构为图 G 中相邻的两顶点间的结构, 并且假设 $v_i^0 (i \in \{1, 2, \dots, n\})$ 与其他顶点都邻接, 图 8(c) 的 DAE 结构处理顶点个数为偶数时的边界, 为了便于控制反应加入顶点 ∞ , 并且假设 ∞ 与其他顶点都邻接. 故规则分子的种类为 $\Theta(n^2 + |E|)$. 由于 $v_i^0 (i \in \{1, 2, \dots, n\})$ 的值不影响提取最大团, 所以可将所有 v_i^0 设计成一样的, 即 $v_1^0 = v_2^0 = \dots = v_n^0$. 故规则分子的种类可以降至 $\Theta(n + |E|)$.

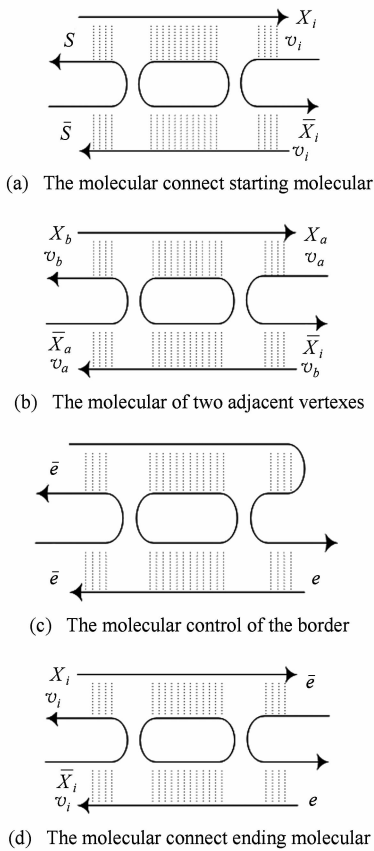


Fig. 8 Abstract structure of the regular molecular.
图 8 规则分子 DAE 抽象结构

2.2.3 检测分子的设计

为了到便于检测并提取最大团,首先设计相应的检测分子提取图的所有团,相应的检测分子的设计如图 9~11 所示.若顶点子集的导出子图为团,该顶点子集的任意两个顶点都连接,则每两个顶点都存在如图 7 所示的规则分子.经过充分退火反应 ∞ 将从与始端 s 相邻逐渐被交换到离终端 e 相邻的位置.由图 9 所示的 DAE 分子就能消除掉 ∞ ,这样终端 e 就向左边移动了一格.然后由图 10 所示的 DAE 分子逐行消去其他顶点,使得终端 e 一直向左移动直到与始端 s 相邻,其中对于顶点 v_i ($i \in \{1, 2, \dots, n\}$) 做特殊处理,留出粘性末端 M 以备下一步提取最大团.最后设计如图 11 所示的带特殊荧光的 DNA 分子连接始端与终端.若最终生成的 DNA 分子格局存在这种带荧光的 DNA 分子,则该格局所

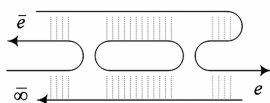


Fig. 9 Abstract structure of DAE when ∞ is eliminated.
图 9 消去 ∞ 的 DAE 分子抽象结构

含顶点集的导出子图就为一个团.由图 9~11 可知用于提取团的检测分子的种类为 $\Theta(n)$.

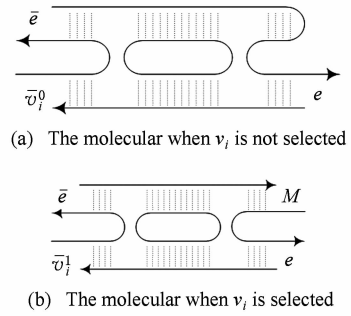


Fig. 10 Abstract structure of DAE when other vertexes are eliminated.

图 10 消去其他顶点的 DAE 分子抽象结构



Fig. 11 DNA molecular with special fluorescence.

图 11 带有特殊荧光的 DNA 分子

为了判定团所表示的顶点数大小,根据上述规则在消去顶点 v_i 时留下的粘性末端 M 设计如图 10 所示的 DAE 分子.加入该类分子充分退火后,若该格局中(如图 12 所示) DAE 分子数最多的就为最大团.为了便于实现该提取操作,引入凝胶电泳技术把 DNA 分子按照大小分离,然后提取分子结构最大的即为最大团.

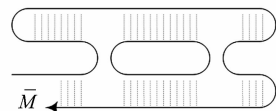


Fig. 12 Abstract structure of DAE for judging the number of vertexes.

图 12 判定团顶点数大小 DAE 分子抽象结构

综上所述,初始分子所需的 DAE 分子种类为 $\Theta(n)$;规则分子所需的 DAE 分子种类为 $\Theta(n + |E|)$;检测分子所需的 DAE 分子种类为 $\Theta(n)$.所以总共需要的 DAE 分子种类为 $\Theta(n + |E|)$,说明设计这些 DAE 分子是可行的.

2.3 算法步骤

算法 1. 基于自组装模型的最大团问题 DNA 算法(maxclique-algorithm).

输入:图 $G = \{V, E\}$,按上述方式设计的 tiles 集 T_1, T_2, T_3, T_4 .其中 T_1 表示利用 PCR 技术大量复制初始分子的试管, T_2 表示利用 PCR 技术大量复制规则分子的试管, T_3 表示利用 PCR 技术大量复制用于提取所有团的检测分子试管, T_4 表示利用

PCR 技术大量复制用于提取最大团的检测分子试管。
输出:图 G 的最大团的顶点集合。

Procedure *maxclique-algorithm* (G, T_1, T_2, T_3, T_4)

- 1) *anneal*(T_1);
- 2) $T = \cup(T_1, T_2)$;
- 3) *anneal*(T);
- 4) $T = \cup(T, T_3)$;
- 5) *anneal*(T);
- 6) $T'_{\text{end}} = \text{extract}(T, d) / *$ 其中 d 为如图 10 所示的具有荧光的检测分子,提取试管 T 中具有荧光的 DNA 分子放入试管 $T'_{\text{end}} / *$;
- 7) $T'_{\text{end}} = \cup(T'_{\text{end}}, T_4)$;
- 8) *anneal*(T'_{end});
- 9) *gel-electrophoresis*(T'_{end}) / * 利用凝胶电泳技术将 T'_{end} 中的 DNA 分子块按照大小进行分离 / *;
- 10) $T_{\text{end}} = \text{extract}(T'_{\text{end}}) / *$ 提取 9) 中最大的 DNA 分子,将其放入试管 $T_{\text{end}} / *$ 。

3 算法性能分析与比较

3.1 算法的正确性证明

引理 1. 若 $\Sigma_1 = \{s, e, v_1, v_2, \dots, v_n\}$, T_1 为按 2.2.1 节设计的初始分子集, $\sigma_1 = 1$, 则形成的自组装系统 $S_1 = \langle T_1, g, \sigma_1 \rangle$. 利用 PCR 聚合酶链式反应对试管 T_1 的各种初始分子迅速大量复制, 充分退火后能够生成表示所有顶点子集的初始格局。

证明. 对于初始分子的设计, 2.2.1 节以顶点个数的奇偶性对初始分子按照两种方式设计, 不妨假设顶点个数为奇数. 由于 $\tau_1 = 1$ 可知两 DAE 分子只需一边的粘性末端互补就能退火声称新的格局. 如图 3 可知始端 DAE 分子为 $\{\text{null}, \text{null}, s, \bar{s}\}$, 与始端相邻的 DAE 分子为 $\{\infty, s, v_1^X, \bar{v}_1\}$, 则这两种 DAE 分子就能退后生成格局 C_1 , 且 $bd_{UR}(A_{C_1}(x_0, y_0)) = s, bd_{UL}(A_{C_1}(x_1, y_0)) = \infty, bd_{UR}(A_{C_1}(x_1, y_0)) = v_1^X, bd_{DR}(A_{C_1}(x_1, y_0)) = \bar{v}_1$. 又由于存在 DAE 分子 $(v_1, v_2^X, v_3^X, \bar{v}_3^X)$ 与格局 C_1 存在一个互补边, 则能够退火形成新的格局 C_2 , 且 $bd_{UR}(A_{C_2}(x_0, y_0)) = s, bd_{UL}(A_{C_2}(x_1, y_0)) = \infty, bd_{UR}(A_{C_2}(x_1, y_0)) = v_1^X, bd_{UL}(A_{C_2}(x_2, y_0)) = v_2^X, bd_{UR}(A_{C_2}(x_2, y_0)) = v_3^X$, 2.2.1 节所示的所有初始分子按照如此规则反应下去则会存在最终格局 $C_{(n+3)/2}$, 且 $bd_{UR}(A_{C_{(n+3)/2}}(x_0, y_0)) = s, bd_{UL}(A_{C_{(n+3)/2}}(x_1, y_0)) = \infty, bd_{UR}(A_{C_{(n+3)/2}}(x_1, y_0)) = v_1^X, bd_{UL}(A_{C_{(n+3)/2}}(x_2, y_0)) = v_2^X, bd_{UR}(A_{C_{(n+3)/2}}(x_2, y_0)) = v_3^X, \dots, bd_{UR}(A_{C_{(n+3)/2}}(x_2, y_0)) = \bar{e}$. 即初始格局为 $(s, \infty, v_1^X, v_2^X, v_3^X, \dots, \bar{e})$, 由于 X_i 等于 1 或者 0, 所以能够生成表示所有顶点子集的初始格局. 证毕。

引理 2. 在初始格局基础上, 若 T_2 为按 2.2.2 节设计的规则分子集, 且 $\sigma_2 = 2$, 则形成自组装系统 $S_2 = \langle T_2, g, \sigma_2 \rangle$. 若对于图 G 对于生成初始格局表示的顶点子集中所有顶点都邻接, 则系统 S_2 会生成格局 C_2 , 其中 ∞ 会逐行向右移动到与终端 e 相邻, 否则不会生成这样的格局。

证明. 初始分子退火生成的初始格局 $(s, \infty, v_1^X, v_2^X, v_3^X, \dots, \bar{e})$, 若该初始格局表示的顶点子集中的所有顶点都邻接, 依据 2.2.2 节则存在规则分子 $(s, \bar{s}, \infty, \bar{\infty}), (v_2^X, v_1^X, \bar{v}_1^X, v_2^X), \dots, (v_n^X, \bar{v}_n^X, \bar{e}, e)$, 并行反应后生成格局 $(s, \infty, v_2^X, v_1^X, v_4^X, v_3^X, \dots, v_n^X, \bar{e})$, 此时则存在规则分子使得并行反应后生成格局 $(s, v_2^X, \infty, v_4^X, \dots, v_{n-3}^X, e)$. 按照上述反应规则, 由于每两个顶点都邻接, $v_i^0 (i \in \{1, 2, \dots, n\})$, ∞ 和其他顶点都邻接, 则反应不会终止, ∞ 逐层向右移动, 直至移动到与终端 e 相邻。

反之, 若存在 v_i^1 与 v_j^1 不邻接, 那么当这两个顶点相邻时则不存在规则分子将其连接, 此时反应终止 ∞ 也不会继续向右移动, 即不能生成上述格局。

证毕。

引理 3. 有 $a_1, a_2, a_3, \dots, a_n$ 个顶点, 按照以下方式循环 n 次比较, 循环次数 i 从 1 逐渐递增到 n : 当 i 为奇数时, 第 $2k (k = \{1, 2, \dots, \lfloor (n+1)/2 \rfloor - 1\})$ 个顶点和第 $2k+1$ 个顶点比较; 当 i 为偶数时, 第 $2k-1 (k = \{1, 2, \dots, \lfloor n/2 \rfloor\})$ 个顶点和第 $2k$ 个顶点比较; 则共需要比较 $C_n^2 = \frac{n(n-1)}{2}$ 次。

证明. 当 n 为偶数时, 则 i 共有 $\frac{n}{2}$ 次为奇数, 此时共比较 $\frac{n}{2} \left(\frac{n}{2} - 1 \right)$ 次, 同时 i 共有 $\frac{n}{2}$ 次为偶数, 则此时共比较 $\frac{n}{2} \times \frac{n}{2}$ 次, 故总共比较 $\frac{n}{2} \left(\frac{n}{2} - 1 \right) + \frac{n}{2} \times \frac{n}{2} = \frac{n(n-1)}{2} = C_n^2$ 次. 同理可知当 n 为奇数时的比较次数亦为 $\frac{n(n-1)}{2} = C_n^2$ 次. 证毕。

定理 1. 若算法 1 中 T_{end} 不为空, 则 T_{end} 中 DNA 分子格局所表示顶点子集的导出子图即为所求图的最大团, 否则该图的任意两个顶点都不邻接。

证明. 若 T_{end} 不为空, 则 T_{end} 中的顶点集是按照引理 3 中的方式进行比较. 由引理 3 可知 T_{end} 中的顶点集共进行了 C_n^2 次比较. 又由引理 1, 2 反映的组装过程可知, 两个顶点比较之后一个顶点向左移动, 另外一个顶点向右移动, 则这两个顶点不会重复比较, 所以每两个顶点都比较过一次. 说明这 n 个顶点两两邻接, 即此顶点子集的导出子图就为此图的一个团. 又由算法 1 的 10) 步可知 T_{end} 中所表示顶点集都规模最大, 所以 T_{end} 中所代表的团即为所求图的最大团. 反之若 T_{end} 为空, 则说明图 G 的任意顶点子集的导出子图都不是团, 即该图的任意两个顶点都不邻接. 证毕.

3.2 算法实例

如图 13, 简单无向图 $G = \{V, E\}$ 且 $V = \{v_1, v_2, v_3, v_4, v_5\}$, $E = \{(v_1, v_2), (v_2, v_3), (v_3, v_4), (v_4, v_5), (v_2, v_5), (v_2, v_4), (v_3, v_5)\}$.

$v_5), (v_2, v_5), (v_2, v_4), (v_3, v_5)\}$.

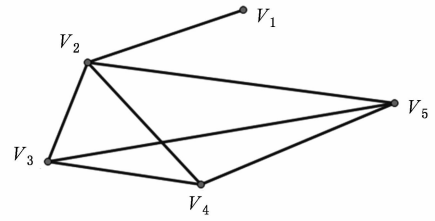


Fig. 13 Simple undirected graph with 5 vertices.
图 13 5 个顶点的简单无向图

由算法 1 可知, 求解图 G 的最大团步骤主要分为 3 个步骤. 具体过程如图 14~16 所示. 对试管 T_1 中的各类初始分子利用 PCR 迅速大量复制, 加入连接酶并使用退火操作使其初始分子充分反映, 则会形成不同的顶点子集初始格局, 如图 14 形成的初始格局为 $\{s, v_1^0, v_2^1, v_3^1, v_4^1, v_5^1, e\}$.

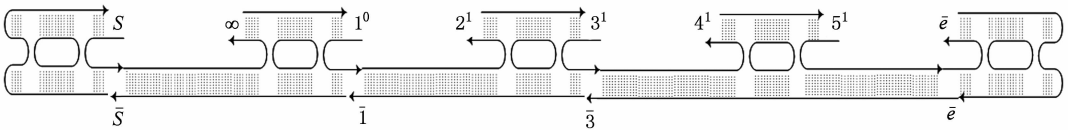


Fig. 14 The initial configuration of $\{v_1^0, v_2^1, v_3^1, v_4^1, v_5^1\}$.
图 14 顶点子集 $\{v_1^0, v_2^1, v_3^1, v_4^1, v_5^1\}$ 形成的初始格局

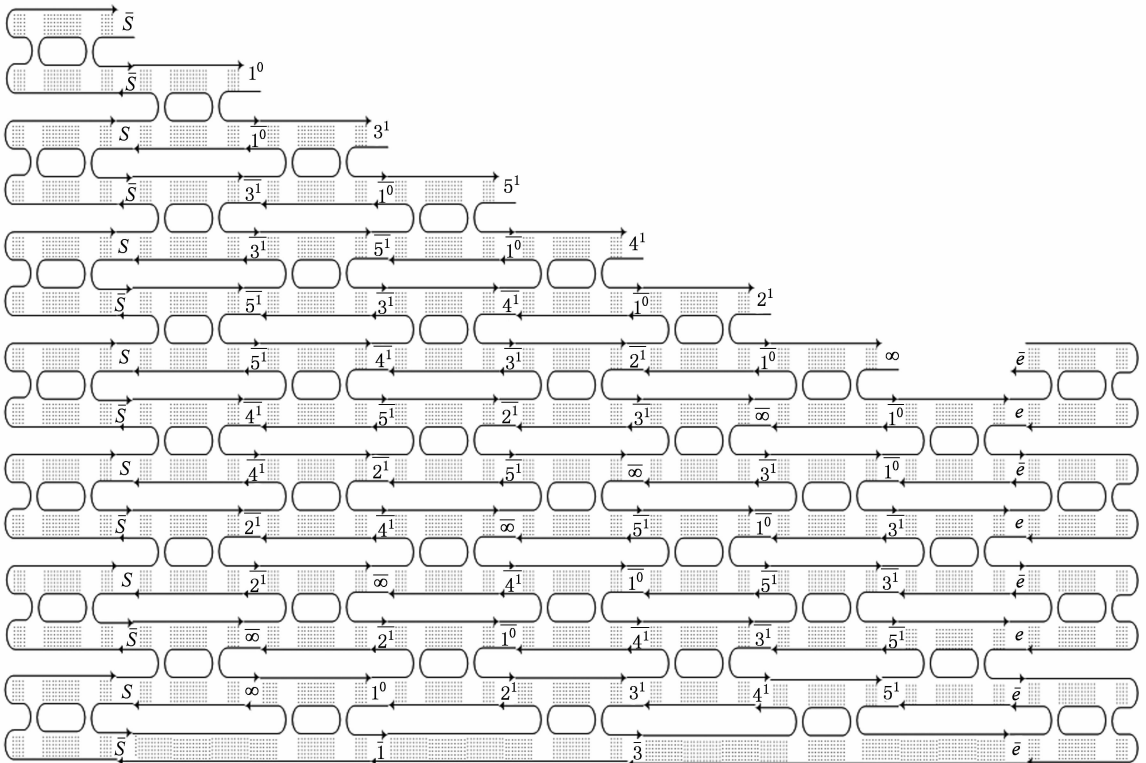


Fig. 15 The configuration after adding regular molecular.
图 15 初始格局 $\{s, v_1^0, v_2^1, v_3^1, v_4^1, v_5^1, e\}$ 加入规则分子后形成的格局

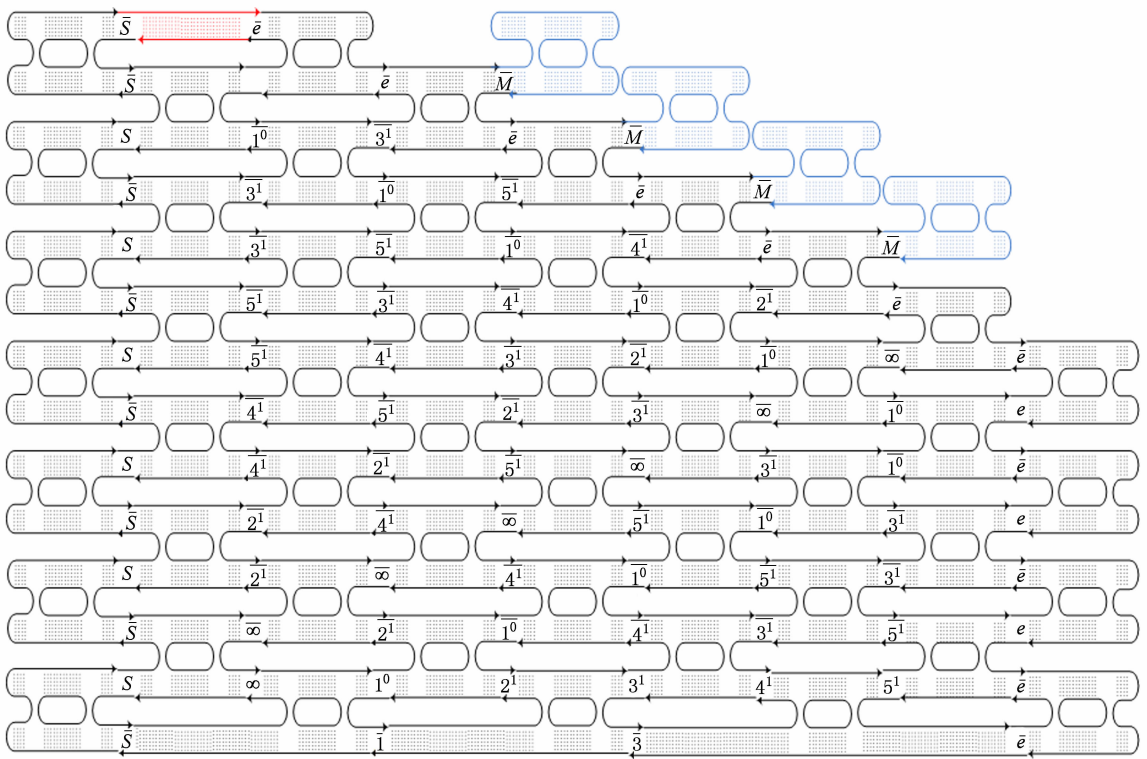


Fig. 16 The final configuration after adding detecting molecular.

图 16 初始格局 $\{s, v_1^0, v_2^1, v_3^1, v_4^1, v_5^1, e\}$ 加入检测分子后形成的最终格局

将试管 T_2 中的各类规则分子利用 PCR 迅速大量复制并与试管 T_1 生成的初始格局合并到试管 T , 之后在试管 T 中加入连接酶, 使用退火操作. 此步骤为加入规则分子, 判断各种初始格局中的任意两个顶点是否都邻接. 如图 15 为加入规则分子用以判断初始格局 $\{s, v_1^0, v_2^1, v_3^1, v_4^1, v_5^1, e\}$ 是否每两个顶点都邻接.

最后加入检测分子, 首先加入提取所有团的检测分子, 提取带有荧光的 DNA 分子. 由定理 1 可知, 如果能够形成如图 16 所示的结构, 则说明该导出子图为团, 即该图中存在有荧光标记的 DNA 分子. 然后加入提取最大团的检测分子, 利用凝胶电泳技术提取最大团.

3.3 算法性能比较

自 1996 年 Adleman 引入 DNA 计算以来, 对于最大团的 DNA 计算算法也有很多研究成果. 1997 年, Ouyang 等人仿效 Adleman 的方法, 提出基于双链 DNA 计算模型求一个图的最大团问题的 DNA 解法^[12]. 2000 年, Head 等人提出基于质粒 DNA 计算模型的算法^[13]. 2010 年, 范月科等人提出基于粘贴模型的算法^[14]. 虽然这些算法能够在多项式时间复杂度内求得给定图的最大团, 但是算法具有较高的生化操作复杂度. 生化实验操作带来的实验误差

较高, 生化实验结果的准确性较低.

本文采用 DNA 自组装模型, 在此模型的基础上提出了最大团问题的 DNA 计算机算法. 算法所需要 tile 的种类为 $\Theta(n + |E|)$, 且算法的生化操作复杂性为 $\Theta(1)$. 与文献[12-15]的算法性能比较, 本算法具有较低的生化操作复杂度, 明显地提高了生化解的准确性, 且具有实验操作性.

4 结 论

最大团问题属于经典的 NP 完全性问题, 除非 $NP=P$, 否则传统的图灵机上不可能存在多项式时间复杂度的有效算法. 自从 Adleman 首次运用 DNA 计算来解决 NP 问题以来, 相继有不少学者提出最大团问题的 DNA 计算机算法. 大多学者都是利用粘贴、剪接模型等, 虽然这些算法能够在多项式时间内求得给定图的最大团, 但是其生化操作具有较高的复杂度. 生化实验操作步骤越多, 相应带来的试验误差也就越大, 从而无法保证生化解的准确性. 本文利用 DNA 自组装模型, 根据 DNA 链形成的 tiles 的自组装特性, 提出了一种新的最大团问题 DNA 算法. 算法所需的 DAE 分子的种类 $\Theta(n + |E|)$, 生化操作复杂度为常数级, 明显地提高了生化解的

准确性,具有良好的实验操作性.

由于本文所提出的算法所需的 DNA 链数为 $O(2^n)$,在求解对于顶点数较多的团问题时还存在困难.因此下一步工作是在保持生化操作复杂度为常数级的前提下,减少算法所需的 DNA 链数.虽然目前还无法确切预知 DNA 生物计算的前景,但注意到近年来 DNA 计算中生化技术所取得的多种进展以及 DNA 计算理论模型的计算完备性证明^[22],具有更少的生化操作复杂度和 DNA 链数的算法研究仍然具有相当意义.

参 考 文 献

- [1] Adleman L M. Molecular computation of solutions to combinatorial problems [J]. *Science*, 1994, 266(11): 1021-1024
- [2] Roweis S, Winfree E, Burgoyne R, et al. A sticker based architecture for DNA computation [J]. *Journal of Computational Biology*, 1998, 5(4): 615-629
- [3] Kari L, Thierrin G. Contextual insertions/deletions and computability [J]. *Information and Computation*, 1996, 131(1): 47-61
- [4] Winfree E, Liu F, Wenzler L A, et al. Design and self-assembly of two-dimensional DNA crystals [J]. *Nature*, 1998, 394(6): 539-544
- [5] Jonoska N, Karl S A, Saito M. Three dimensional DNA structures in computing [J]. *Biosystems*, 1999, 52(1/2/3): 143-153
- [6] Head T, Rozenberg G, Bladergroen R S, et al. Computing with DNA by operating on plasmids [J]. *BioSystems*, 2000, 57(2): 87-93
- [7] Wang H. Proving theorems by pattern recognition I [J]. *Communications of the ACM*, 1960, 3(4): 220-234
- [8] Fu T J, Seeman N C. DNA double crossover molecules [J]. *Biochemistry*, 1993, 32(13): 3211-3220
- [9] Seeman N C. DNA nanotechnology: Novel DNA constructions [J]. *Annual Review of Biophysics and Biomolecular Structure*, 1998, 27(1): 225-248
- [10] Winfree E. *Algorithmic self-assembly of DNA* [D]. Pasadena: California Institute of Technology, 1998
- [11] Yu H, Tao Y, Min S, et al. Hierarchical self-assembly of DNA into symmetric supramolecular polyhedral [J]. *Nature*, 2008, 452(3): 198-201
- [12] Ouyang Q, Kaplan P D, Liu S M, et al. DNA solution of the maximal clique problem [J]. *Science*, 1997, 278(10): 446-449
- [13] Head T, Rozenberg G, Bladergroen R B, et al. Computing with DNA by operating on plasmids [J]. *Biosystems*, 2000, 57(2): 87-93
- [14] Fan Yueke, Qiang Xiaoli, Xu Jin. Sticker model for maximum clique problem and maximum independent set [J]. *Chinese Journal of Computers*, 2010, 33(2): 305-310 (in Chinese)
(范月科, 强小利, 许进. 图的最大团与最大独立集粘贴 DNA 计算模型 [J]. *计算机学报*, 2010, 33(2): 305-310)
- [15] Li Kenli, Zhou Xu, Zhou Shuting. Improved volume molecular solutions for the maximum clique problem on DNA-based supercomputing [J]. *Chinese Journal of Computers*, 2008, 31(12): 2173-2181 (in Chinese)
(李肯立, 周旭, 邹舒婷. 一种改进的最大团问题 DNA 计算算法 [J]. *计算机学报*, 2008, 31(12): 2173-2181)
- [16] Yang Jing, Zhang Cheng. Progress and difficulty in DNA self-assembly technology [J]. *Chinese Journal of Computers*, 2008, 31(12): 2138-2148 (in Chinese)
(杨静, 张成. DNA 自组装技术的研究进展及难点 [J]. *计算机学报*, 2008, 31(12): 2138-2148)
- [17] Fu T J, Seeman N C. DNA double-crossover molecules [J]. *Biochemistry*, 1993, 32(13): 3211-3220
- [18] Brun Y. Solving NP-complete problems in the tile assembly model [J]. *Theoretical Computer Science*, 2008, 395(1): 31-46
- [19] Xu J, Bao Z. Neural networks and graph theory [J]. *Science in China Series F: Information Science*, 2007, 45(1): 1-24
- [20] Dirk F, Anthony R C, Richard J L, et al. Molecular computation: RNA solutions to chess problems [J]. *Biochemistry*, 2000, 39(4): 1385-1389
- [21] Holland J H. *Adaptation in Nature and Artificial Systems* [M]. Cambridge, MA: MIT Press, 1992
- [22] Matteo C, Peter L. Computing by observing: Simple systems and simple observers [J]. *Theoretical Computer Science*, 2011, 412(1/2): 113-123



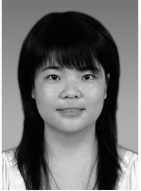
Li Kenli, born in 1971. PhD and professor. Senior member of China Computer Federation. His main research interests include parallel computing, molecular computing.



Luo Xing, born in 1987. Master candidate. His main research interests include DNA Computing.



Wu Fan, born in 1973. PhD candidate. His main research interests include DNA computing.



Zhou Xu, born in 1983. Master and assistant professor. Her main research interests include parallel computing and DNA computing.



Huang Xin, born in 1984. PhD candidate. His main research interests include parallel and disturbed computing.

《计算机研究与发展》征订启事

《计算机研究与发展》(Journal of Computer Research and Development)是中国科学院计算技术研究所和中国计算机学会联合主办、科学出版社出版的学术性刊物,中国计算机学会会刊. 主要刊登计算机科学技术领域高水平的学术论文、最新科研成果和重大应用成果. 读者对象为从事计算机研究与开发的研究人员、工程技术人员、各大专院校计算机相关专业的师生以及高新企业研发人员等.

《计算机研究与发展》于 1958 年创刊,是我国第一个计算机刊物,现已成为我国计算机领域权威性的学术期刊之一. 并历次被评为我国计算机类核心期刊,多次被评为“中国百种杰出学术期刊”. 此外,还被《中国学术期刊文摘》、《中国科学引文索引》、“中国科学引文数据库”、“中国科技论文统计源数据库”、美国工程索引(EI)检索系统、日本《科学技术文献速报》、俄罗斯《文摘杂志》、英国《科学文摘》(SA)等国内外重要检索机构收录.

国内邮发代号:2-654;国外发行代号:M603

国内统一连续出版物号:CN11-1777/TP

国际标准连续出版物号:ISSN1000-1239

联系方式:

100190 北京中关村科学院南路 6 号《计算机研究与发展》编辑部

电话: +86(10)62620696(兼传真); +86(10)62600350

Email:crad@ict.ac.cn

http://crad.ict.ac.cn