



自组装 DNA 链置换分子逻辑计算模型

张成^{①†}, 马丽娜^{②†}, 董亚非^{②*}, 杨静^③, 许进^{①*}

① 北京大学信息科学技术学院, 高可信度软件技术教育部重点实验室, 北京 100871;

② 陕西师范大学生命科学学院, 西安 710062;

③ 华北电力大学控制与计算机工程学院, 北京 102206

* 联系人, E-mail: dongyf@snnu.edu.cn; jxu@pku.edu.cn

† 同等贡献

2012-06-01 收稿, 2012-09-04 接受

国家自然科学基金重大仪器专项(61127005)、国家自然科学基金重点项目(61133010, 61033003)、国家自然科学基金国际重大合作项目(60910002)、国家自然科学基金面上项目(61143003)、教育部博士点基金(20110001130016)、中国博士后特别资助项目(201104018)和博士后基金(20100480128, 2011M500197)资助

摘要 将 DNA 自组装与链置换技术相结合, 构建了 DNA 分子逻辑计算模型. 通过输入 DNA 信号链, 经特异性识别和链置换, 自组装初始结构发生变化, 并释放特定信号分子或结构作为输出. 计算模型能通过 DNA 自组装结构电泳迁移率的改变输出计算结果. 计算系统在室温下可自动触发、混合输入信号以及并行置换. 另外, 本模型中引入了单极和双极两种置换模式设计, 其还具有并行信息处理和模块化组装等特点. 最后经实验验证, 通过输入特异性 DNA 分子链置换, 该计算模型能正确输出逻辑计算结果.

关键词

DNA 链置换
DNA 自组装
分子逻辑计算
分子智能
模块化设计

近年来, 随着“摩尔定律”的瓶颈逼近, 面对海量信息处理, 传统电子计算压力倍增, 人们将目光逐渐转移到新型计算领域. 基于全新模式的量子分子计算脱颖而出, 科学家试图利用纳米材料和技术, 来实现超大规模信息处理. 特别是 DNA 计算领域, 其融合了信息、生物、纳米等多个学科, 成为有广泛应用前景的热点研究领域^[1-5]. 由于 DNA 分子的天然特性: 特异性、高并行性和微小性等, 在海量信息存储和处理过程中, 可以高容量保存以及并行操作, 具有显著优势. 这使得 DNA 计算极有可能成为未来计算的实现途径. 自 20 世纪末至今, DNA 计算在理论、设计以及技术实现等方面都取得了较大的进展, 在信息处理、分子智能、分子密码、纳米机器人和纳电子学等多个领域得到深入的研究^[6-11].

DNA 计算发展至今, 融合了多种分子操作技术: PCR 技术、荧光标记、链置换、纳米颗粒和自组装技术等. 其中, 结合荧光标记的 DNA 链置换技术发展迅速. 科学家已经利用该技术构建了多种分子计

算模型^[12-15]. 值得注意的是, 2011 年 *Science* 和 *Nature* 报道了加州理工学院计算机系的 Winfree 教授研究组利用 DNA 链置换技术实现简单的开平方运算和神经网络计算, 引起了信息计算领域的极大关注^[4,5]. DNA 分子链置换技术不仅能与荧光检测、DNA 自组装结合, 而且还可以进一步与纳米颗粒、量子点、蛋白质等结合, 从而促进了包括并行计算、密码集成、纳电子学等研究领域的共同发展.

本文将 DNA 自组装与链置换技术相结合, 构建了 DNA 链置换触发的逻辑计算模型. 总体上, 该运算模型可分为两个部分: DNA 自组装结构和链置换“控制器”. 计算时, 先利用 DNA 杂交形成各类分子自组装结构, 构建基本的计算单元. 再通过分别输入特异性的 DNA 分子信号, 对基本单元发出指令, 释放特定输出分子信号. 本模型能在室温下自发进行, 且具有高度准确性(结果可通过多种电泳条带检测). 结合单极和双极置换设计, 模型具有并行信息处理和模块化组装等特点. 针对特异输入信号, 通过比较

DNA 自组装枝杈结构电泳迁移率变化, 该计算模型能准确进行逻辑运算.

1 DNA 自组装和链置换

1.1 DNA 自组装

DNA 自组装是指以 DNA 分子为基本材料, 通过分子间的碱基氢键、范德华力等微作用力自发缔结成结构稳定的聚集体或超分子结构. 整个自组装是一个由简单到复杂、从杂乱无章到井然有序的自发过程, 不需要人为干预. 20 世纪 60 年代, 提出 DNA 自组装概念, 设想利用 DNA 单链组装成瓦片结构. 然而直至 20 世纪 90 年代, DNA 自组装技术才变为现实. 这一开创性的工作, 主要由 Seeman 教授研究组完成的. 他们通过多条短链 DNA 分子杂交, 获得了四边形、圆环、绳结状等复杂结构^[16~18]. Seeman 教授本人也因此获得了 2010 年世界 Kavli 纳米领域奖. DNA 自组装的另一个思路是利用较长的单链 M13 噬菌体 DNA. 这种方法成本低、结构多样以及产量高, 非常适合于大规模操作. 2006 年, Rothemund 借助 M13 DNA 链和短的寡核苷酸链, 折叠获得平面自组装结构^[19]. 随着自组装技术的成熟, 近年来多种立体复杂自组装结构也相继被构建^[20~24].

1.2 DNA 链置换

DNA 链置换是利用分子杂交系统自由能趋向于稳定的特点, 通过改变输入信号分子的长度和序列, 控制或诱发下游链置换反应^[4,5]. 直观上理解, 就是用较长的双链杂交取代较短杂交区域的过程. DNA 链置换(DNA Strand Branch Migration)基本过程见图 1. 首先, 两条部分互补的 DNA 链结合在一起(链 A 长度大于链 B). 在室温下, 待加入链 C(其序列完全与链 A 互补), 特异性识别区域先结合 A 链单链部分. 为了达到最稳态, 再逐渐占领 A 链与 B 链的结合位点, 最后完全置换释放出 B 链. 由于该过程能并行启动, 具有高度特异性识别位点, 且能够进行多层次嵌套触发, 因此近年来发展迅速, 已经成为分子计算领域的前沿研究热点^[12~14].

2 实验方法

2.1 链置换逻辑计算模型结构和序列设计

逻辑运算又称为布尔运算. 近代科学家布尔用

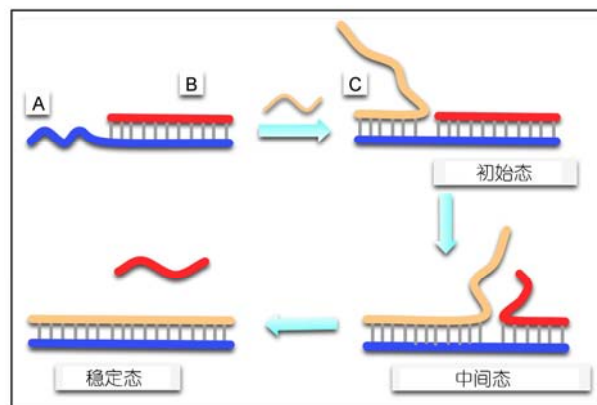


图 1 DNA 链置换基本过程

数学方法研究逻辑计算, 并成功地建立了逻辑演算方法. 20 世纪 30 年代在电路系统领域得到广泛应用. 由于传统的逻辑电路组成基本只有两个状态, 即电流的开与断, 这与逻辑计算的输入输出非常符合. 随着近代微电子技术的发展, 虽然出现各种复杂超大系统, 但它们的最基本运算单元都是利用逻辑计算原理实现的. 2011 年, *Nature* 对 DNA 分子实现的逻辑运算器做了专题介绍^[25]. 因此, 无论是对于计算机科学, 还是其他相关交叉领域而言, 利用自然分子模拟基本逻辑计算都具有一定理论和应用意义.

本研究中, 利用 DNA 链置换进行实现与门(AND)和或门(OR)基本逻辑运算. 其中, 利用单极和双极链置换, 分别设计构建“AND”和“OR”门, 并通过实验进行验证.

(1) 链置换逻辑计算模型设计. 构建链置换逻辑计算模型主要包括两方面的工作: 一是设计 DNA 分子自组装结构, 识别区域设计和非信号区的结构包埋等; 二是整个系统 DNA 序列设计, 着重考虑分子内部和链间互补错配. 另外, 由于运算方式和原理不同, 需要设计相应的 DNA 自组装结构. 然后再通过编码软件及限制条件, 筛选出合适的 DNA 序列. 设计过程中要重点注意避免碱基错配、DNA 茎环结构的形成和多聚体等问题.

在自组装“AND”模块结构中, 为了便于模块化设计和控制链置换反应, 其主体结构是由两种链杂交而成的: 主体结构链(A)和信号识别链(ab 和 cd)(图 2). “AND”模块采用单极置换模式, 即输入信号链(ab-1 或 cd-1)只能与单侧信号识别链 ab 或 cd 相作用. 其中, 主体结构链 A 由 20 bp 碱基组成, 分为左右两个各长 10 bp 的片段, 其作用为连接两端的信号识别

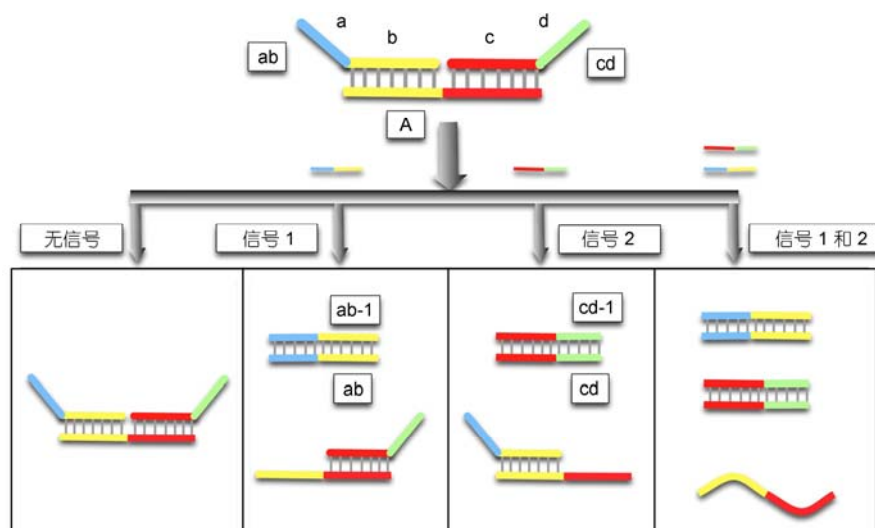


图 2 自组装“AND”模块置换原理图

链 ab 和 cd . 而信号识别链 ab 和 cd 又可分为两个部分: 结构区域 b 和 c , 以及特异性识别位点 a 和 d . 在这里, 对信号识别链的结构区域和识别位点进行区别设计. 结构区域因其承担框架结构, 不能轻易解构, 因此设计链长 10 bp. 而特殊识别位点是进行链置换的关键部位, 既需要实现结构功能, 又要完成特异性识别, 因此设计链长约 6~7 bp. 如图所示, 颜色相同的 DNA 区域为互补序列. 经置换后, 形成了一系列新的 DNA 自组装枝杈结构.

类似的, 自组装“OR”模块结构也采用链置换反

应模式实现计算. 自组装设计由主体结构链(B)和多个信号识别区域(e, f, g, h)组成. 但不同的是, “OR”模块使用了双极同步链置换方式(图3). 即利用不同识别区域的相同识别位点, 对其进行同步置换. 双极同步链置换有利于实现模块化设计, 即在仅改变部分链序列的基础上, 就可以实现逻辑关系的计算. 另外, “OR”模块中还采用了混合输入信号作为单种输入, 如信号 3 和 4 分别为输入链(ef-1, eg-1)和(fh-1, gh-1). 置换后, 形成了新的部分互补双链 DNA 和单链 DNA 分子.

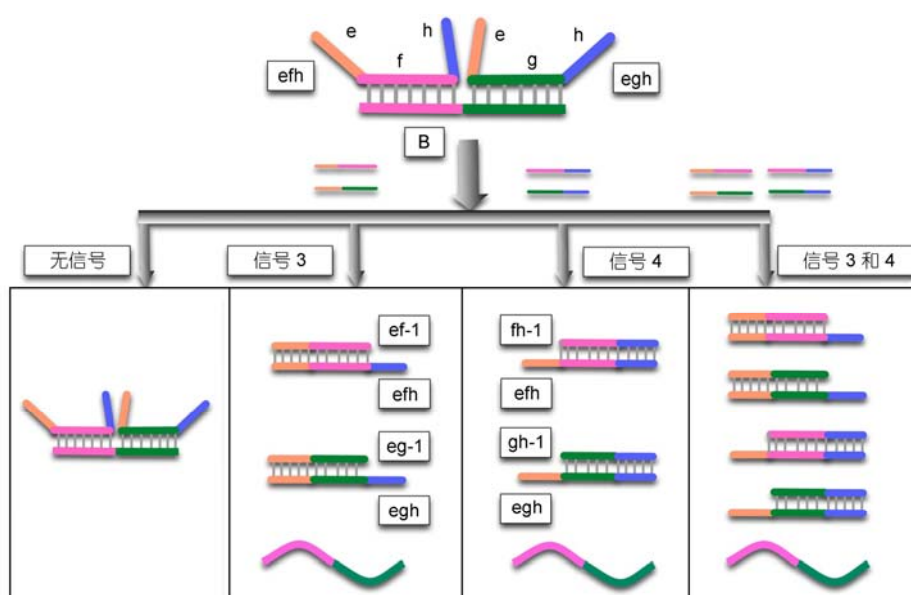


图 3 自组装“OR”模块置换原理图

(2) 输入信号分子链触发链置换. 向不同逻辑计算模块中加入相应的输入信号, 通过特异性识别位点相互作用, 对应的信号分子输出链被释放(见图 2 和 3).

(3) 检测计算结果. 通过聚丙烯酰胺凝胶电泳 (PAGE) 检测出逻辑计算相应的结果.

2.2 实验方法和材料

(1) 实验方法. 逻辑门通过加入各种结构 DNA 分子后, 经缓慢退火杂交得到. 具体方法如下所述: 构建 DNA 逻辑计算模型时, 将等摩尔浓度的各种 DNA 链加入到体系中(终浓度 1 pmol/μL, TAE/Mg2+ 缓冲液). 将混合液体置于如下退火条件: 95℃ 4 min, 65℃ 30 min, 50℃ 30 min, 37℃ 30 min, 22℃ 30 min. 此步骤后, 可进行电泳进行检测. 实验中, DNA 链置换在 1×PBS 缓冲液中实现. 加入的 DNA 链和运算逻辑模块混合后, 在室温下反应 6 h 以上, 即可进行 PAGE 电泳结果检测. 链置换反应产物可以置于 4℃ 下保存.

(2) 实验试剂及其配制. 实验中所使用的 DNA 链购自上海生工, 都经过 PAGE 电泳纯化. 实验中的 DNA 分子编码序列设计(见表 1).

Tris 碱、Na₂EDTA、溴化乙锭(EB)、丙烯酰胺、亚甲双丙烯酰胺、过硫酸铵、四甲基乙二胺(TEMED) 和 Stain all 购自 Sigma 公司产品; DNA Marker(DNA marker 20 bp ladder, 50 bp ladder)、溴酚兰购自 TaKaRa 公司; 10×PBS 缓冲液购自北京索莱宝公司; TAE/ Mg²⁺缓冲液: 0.04 mol/L Tris acetate, 1 mmol/L EDTA, 12.5 mmol/L Mg acetate, 调整 pH 为 8.3; 丙烯酰胺母液: 配制 500 mL 浓度为 45%的丙烯酰胺母液, 称取 217 g 丙烯酰胺和 8 g 亚甲双丙烯酰胺, 37℃ 下促溶解, 加入去离子水定容至 500 mL.

3 实验结果

本实验中, 利用 PAGE 电泳进行结果检测. 由于 DNA 自组装结构发生变化, 会导致高分子复合体电泳速度发生变化, 因此通过分析泳道 DNA 条带的特异变化, 即可对计算结果进行输出.

在“AND”逻辑计算模块中, 分四种情况进行输入: 无信号、信号 1、信号 2 以及共同加入信号 1 和 2. (1) 当没有信号输入时, 易知自组装结构没有变化. (2) 当信号 1(ab-1)加入时, 特异性识别区域 a 相互作用, 输入链 ab-1 将 ab 链完全置换. (3) 当信号 2(cd-1)加入时, 特异性识别区域 d 相互作用, 输入链 cd-1 将 cd 链完全置换. (4) 当同时加入信号 1 和 2 时, 特异性识别区域 a 和 d 同时作用, 此时彻底释放 DNA 链 A. 从结构看, 只有同时加入信号 1 和 2, 才能够导致自组装结构完全解体. 从电泳结果图 4 看, 泳道 1~5 显示了各个不同自组装体的电泳条带, 泳道 6,8,9 和 10 显示了计算结果. 计算结果可以从电泳条带部位进行判读, 以橘红色箭头标出(标记部位代表单链 A 的电泳位置). 从电泳结果可知, 只有泳道 10 置换释放出了链 A, 即输出值为真(信号 1 和 2 同时输入). 而在泳道 6,8 和 9 同样的位置, 都没有 A 链被释放, 即输出值为假. 另外, 蓝色箭头标记部位, 其代表未加入信号时, 保持的完整自组装体电泳条带(泳道 6).

图 4 中, 泳道 6 为逻辑门初始状态自组装体. 一方面, 通过对比泳道 3 的 A 链电泳位置, 发现泳道 8 和 9 的对应位置并无释放 A 电泳条带. 而双聚体的电泳位置则出现在泳道 3 中 A 条带靠下的位置. 另一方面, 同时加入信号 1 和 2 后, 泳道 10 中产生了新的电泳条带(橘红箭头标记). 通过与泳道 3 比较, 发现其位置与 A 链相当. 并且泳道 10 位置靠下的条带(双聚体 ab/ab-1 和 cd/cd-1)比同泳道中的 A 链条带位置明显滞后. 因此, 可以推断新产生的电泳条带就是释放

表 1 实验中所使用的 DNA 序列

名称	序列(5'~3')	名称	序列(5'~3')
A	TCTGGCACTATGACAAGCGA	efh	CTAATCACCTGCTTCGGAAGTGG
ab	TAGTGCCAGAGGTATCC	ef-1	CCAGTTCCGAAGCAGGG
ab-1	GGATACCTCTGGCACTA	eg-1	CCAGTTCAATGCGTATG
cd	ATTCATCTCGCTTGTC A	egh	CTAATCACATACGCATTGAACTGG
cd-1	TGACAAGCGAGATGAAT	fh-1	CGAAGCAGGGTGATTAG
B	CGAAGCAGGGAATGCGTATG	gh-1	AATGCGTATGTGATTAG

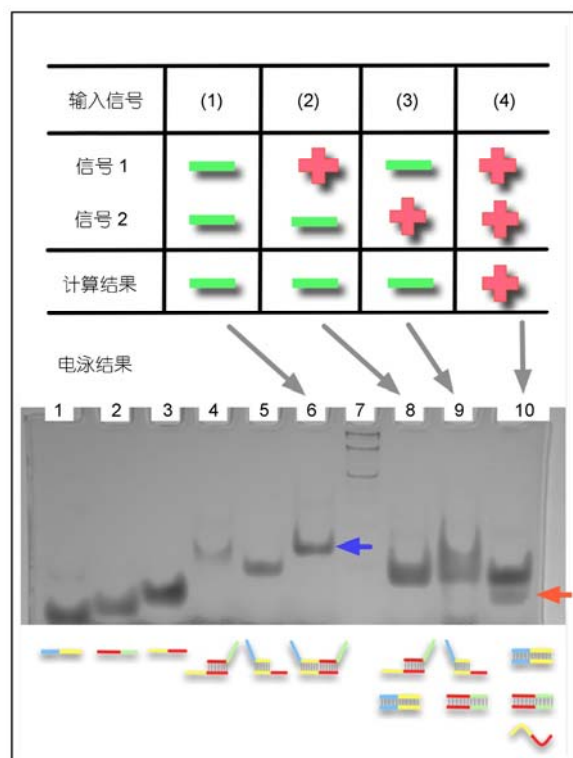


图 4 “AND”逻辑模块计算结果(泳道 7 为 DNA Marker)

的 A 链。

类似的，在“OR”逻辑计算模块中，也分四种情况进行输入：无信号、信号 3、信号 4 以及共同加入信号 3 和 4。但与“AND”计算模块不同的是，每种信号都由两种 DNA 链混合而成，同步对特异性识别位点进行置换。(1) 当没有信号输入时，整体自组装结构没有变化。(2) 当信号 3(ef-1 和 eg-1 链)加入时，特异性识别区域 e 相互作用，输入链 ef-1 和 eg-1 同时对 efh 和 egh 链进行置换。(3) 当信号 4(fh-1 和 gh-1 链)加入时，特异性识别区域 h 相互作用，输入的混合链可将 efh 和 egh 链完全置换。(4) 当同时加入信号 3 和 4 时，特异性识别区域 e 和 h 同时作用，此时完整的初始自组装结构解体。从结构看，只要加入信号 3 或 4，都能够导致自组装结构完全解体。从电泳结果图 5 看，泳道 5,6,7 和 8 显示了计算结果。当没有信号链输入时，自组装结构保持初始状态而无变化，输出值为假，以蓝色箭头标出(泳道 5)。而当加入任意信号后，链置换就会发生，从而导致自组装结构解体。在电泳中表现为电泳速度加快，橘红色箭头标记部位代表解体后的电泳位置。加入任意信号 3 和 4，电泳结果(泳道 6, 7 和 8)都显示出初始自组装体条

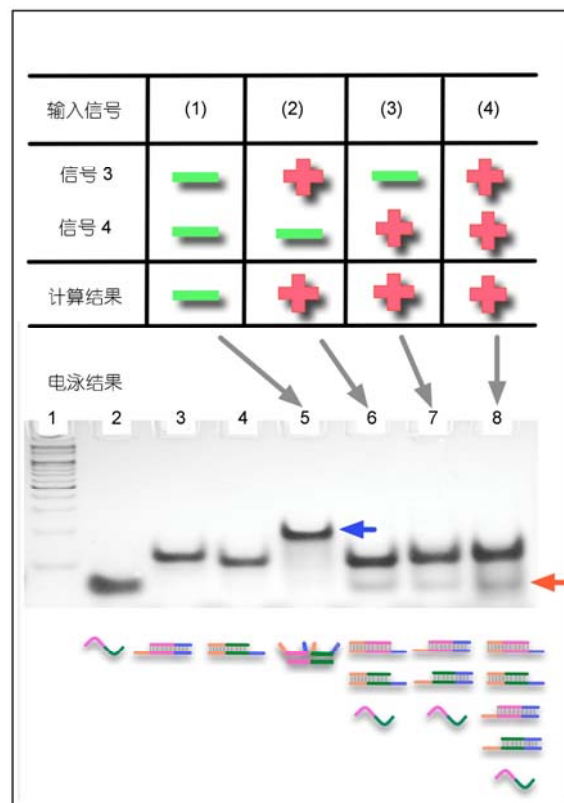


图 5 “OR”逻辑模块计算结果(泳道 1 为 DNA Marker)

带消失，而在双聚体和单链条带(橘红色箭头)位置产生了新的电泳条带。因此，可以推断在单链相应位置处产生的新条带即为输出链。从电泳结果可知，泳道 6,7 和 8 都置换解体了原有的自组装结构，即输出值为真(信号 3 或 4 任意输入)。实验结果与预期符合。

4 讨论

本文利用 DNA 自组装与链置换技术，构建了 DNA 链置换控制触发的逻辑计算模型。计算模型主要由两部分组成：自组装结构模块和链置换触发模块。通过输入 DNA 信号链，经特异性识别和链置换，自组装初始结构发生变化，并释放特定信号分子作为输出。该模型具有如下特点：(1) 室温下自发反应。因为整个链置换反应体系是在室温下进行的，因此不会引起其他自组装结构热变性变化，从而减少了实验误差。(2) 单极和双极置换模式。由于采用了单极和双极置换模式，本模型具有并行信息处理和模块化组装等特点。这样设计主要有如下优点：第一，设计信号链数据库的模块化。仅通过更换不同的识别位点就可以实现新的信号识别，便于大规模程

化设计(在这个过程中,主体结构链是无需进行改变的);第二,同一信号链可以同时触发不同的计算模块.这有利于多种计算模块相互结合和搭建,以及运算的并行操控.第三,输入信号不仅可以是一种链,还可以将多种混合信号链作为一种信号进行输入,可通过组合增加有限的信号数量.(3) 实验结果的高准确判读模式.由于自组装结构变化会导致各类 DNA 复合分子电泳率变化,所以计算输出结果可以综合考虑不同电泳条带的位置进行判读,如解体或新产生电泳条带.该特点有助于提高计算模型的准确率.通过实验验证,针对特异输入信号,该计算模型能准确进行逻辑运算.

同时,在模型构建过程中,该模型还存在以下几个难点和不足.(1) 复杂系统的 DNA 序列设计.该模型采用了双极链置换,对 DNA 序列设计提出了更高的要求.以往设计中,要尽量减少 DNA 序列相似性,而在本实验中,由于需要双极链置换,这就要求两个置换链必须具有相同的特异性识别位点(如输入信号链 ef-1 和 eg-1).对于给定部分序列的 DNA 分子,其设计难度更大,错配现象也更多.(2) 识别混合输入信号链.由于混合输入信号链,使得体系复杂度增加.因此还需要进一步摸索,通过体系浓度调节和控制其他反应条件,有效提高混合信号的识别特异性.(3)

检测方法相对单一和简单.本实验主要通过自组装结构电泳迁移率变化进行结果检测.在日后的研究中,应考虑结合荧光标记、原子力显微镜、生物酶和纳米颗粒等技术,实现综合性多类别结果检测^[26-28].

(4) 释放信号链亮度偏低.其原因可能有如下两个.第一,双链 DNA 染色能力较强.由于染色剂容易插入双链 DNA 之间,因此双链 DNA 比单链 DNA 更容易染色.第二,链置换效率不高.释放链亮度较低,可能是由链置换效率较低引起的.但是,从另一个角度看,初始自组装体在原有电泳位置已经完全消失,而转化为双聚体和单链.这基本可以证明初始自组装体大部分都进行了链置换反应.因此,释放链亮度偏低可能是由于不同 DNA 自组装体的染色能力造成的.在随后的研究中,力争结合其他前沿技术,进一步对该计算模型进行提高和完善.

虽然本文举例求解的规模较小,尚不能解决实际问题,但模型实验的成功操作还是能够说明:DNA 自组装结构和链置换技术在分子计算领域具有的潜在能力.另外,该模型也为相关领域如计算机科学、密码学、纳米智能和医学等提供了新的思路和结合点.随着分子计算理论和实验的不断进步,DNA 计算在不久的将来有望真正地走入应用领域,为未来大规模复杂信息处理发挥作用.

参考文献

- 1 Adleman L. Molecular computation of solution to combinatorial problems. *Science*, 1994, 66: 1021-1024
- 2 Ouyang Q, Kaplan P D, Liu S, et al. DNA solution of the maximal clique problem. *Science*, 1997, 278: 446-449
- 3 Mao C D, LaBean T H, Reif J H. Logical computation using algorithmic self-assembly of DNA triple-crossover molecules. *Nature*, 2000, 28: 493-496
- 4 Qian L L, Winfree E. Scaling up digital circuit computation with DNA strand displacement cascades. *Science*, 2011, 332: 1196-1201
- 5 Qian L L, Winfree E, Bruck J. Neural network computation with DNA strand displacement cascades. *Nature*, 2011, 475: 368-372
- 6 Pei R J, Matamoros E, Liu M H, et al. Training a molecular automaton to play a game. *Nat Nanotechnol*, 2010, 5: 773-777
- 7 Gu H Z, Chao J, Xiao S J, et al. Molecular robots guided by prescriptive landscapes. *Nature*, 2010, 465: 202-205
- 8 Rinaudo K, Bleris L, Maddamsetti R. A universal RNAi-based logic evaluator that operates in mammalian cells. *Nat Biotechnol*, 2007, 25: 795-801
- 9 Maune H T, Han S P, Barish R D, et al. Self-assembly of carbon nanotubes into two-dimensional geometries using DNA origami templates. *Nat Nanotechnol*, 2010, 5: 61-66
- 10 Morrow T J, Li M W, Kim J, et al. Programmed assembly of DNA-coated nanowire devices. *Science*, 2009, 323: 352-353
- 11 Clelland C T, Risca V, Bancroft C. Hiding messages in DNA microdots. *Nature*, 1999, 399: 533-534
- 12 Seelig G, Soloveichik D, Zhang Y D, et al. Enzyme-free nucleic acid logic circuits. *Science*, 2006, 314: 1585-1588
- 13 Shin J S, Pierce N A. Rewritable memory by controllable nanopatterning of DNA. *Nano Lett*, 2004, 4: 905-909
- 14 Zhang D Y, Turberfield A J, Yurke B, et al. Engineering entropy-driven reactions and networks catalyzed by DNA. *Science*, 2007, 318: 1121-1125

- 15 Frezza B M, Cockroft S L, Ghadiri M R. Modular multi-level circuits from immobilized DNA-based logic gates. *J Am Chem Soc*, 2007, 129: 14875–14879
- 16 Winfree E. Design and self-assembly of two-dimensional DNA crystals. *Nature*, 1998, 394: 539–544
- 17 LaBean T H, Reif J H, Seeman N C, et al. Construction, analysis, ligation, and self assembly of DNA triple crossover complexes. *J Am Chem Soc*, 2000, 122: 1848–1860
- 18 Carbone A, Seeman N C. Circuits and programmable self-assembling DNA structures. *Proc Natl Acad Sci USA*, 2002, 99: 12577–12582
- 19 Rothmund P W. Folding DNA to create nanoscale shapes and patterns. *Nature*, 2006, 440: 297–302
- 20 Goodman R P, Schaap I A T, Tardin C F, et al. Rapid chiral assembly of rigid DNA building blocks for molecular nanofabrication. *Science*, 2005, 310: 1661–1665
- 21 Zhang C, Su M, He Y, et al. Conformational flexibility facilitates self-assembly of complex DNA nanostructures. *Proc Natl Acad Sci USA*, 2008, 105: 10665–10669
- 22 Aldaye F A, Lo P K, Karam P, et al. Modular construction of DNA nanotubes of tunable geometry and single- or double-stranded character. *Nat Nanotechnol*, 2009, 4: 349–352
- 23 Han D R, Pal S, Nangreave J, et al. DNA origami with complex curvatures in three-dimensional space. *Science*, 2011, 332: 342–346
- 24 Ackermann D, Schmidt T L, Hannam J S, et al. A double-stranded DNA rotaxane. *Nat Nanotechnol*, 2010, 5: 436–442
- 25 Thomas C. DNA as a logic operator. *Nature*, 2011, 469: 45–46
- 26 Liu Q H, Wang L M, Frutos A G, et al. DNA computing on surfaces. *Nature*, 2000, 403: 175–179
- 27 Benenson Y, Elizur T P, Adar R, et al. Programmable and autonomous computing machine made of biomolecules. *Nature*, 2001, 411: 430–434
- 28 Benenson Y, Gil B, Dor U B, et al. An autonomous molecular computer for logical control of gene expression. *Nature*, 2004, 429: 423–429