文章编号:2095-476X(2014)01-0022-07

自组装 DNA 计算仿真软件及其应用综述

王子成^{1,2}, 叶盟盟¹, 侯贺伟¹, 韩栋¹, 孙中华¹, 崔光照^{1,2}

(1. 郑州轻工业学院 电气信息工程学院,河南 郑州 450002;2. 河南省信息化电器重点实验室,河南 郑州 450002)

摘要:综述了用于自组装 DNA 计算仿真软件 caDNAno 工具、Visual DSD 工具和 NUPACK 工具的发展现状,并以基于 caDNAno 软件平台设计郑州轻工业学院校徽二维 DNA 自组装结构模型、采用 Visual DSD 工具设计用于素数判断的 DNA 分子计算模型并进行相关仿真实验为例,对自组装 DNA 计 算仿真软件的应用进行可行性论证与探讨,指出:随着自组装 DNA 计算研究的深入,应用新的生物 技术开发更为高效实用的软件将是未来的研究方向.

关键词:DNA 计算;自组装;仿真软件

中图分类号:TM711 文献标志码:A DOI:10.3969/j.issn.2095-476X.2014.01.004

Review of simulation computation of DNA self-assembly system and the related application research

WANG Zi-cheng^{1,2}, YE Meng-meng¹, HOU He-wei¹, HAN Dong¹, SUN Zhong-hua¹, CUI Guang-zhao^{1,2}

(1. College of Electric and Information Engineering, Zhengzhou University of Light Industry, Zhengzhou 450002, China;
2. He'nan Key Laborary of Information-based Electrical Appliances, Zhengzhou 450002, China)

Abstract:Some existing DNA computing softwares, such as caDNAno tools, Visual DSD tools and NU-PACK tools, were reviewed. Based on the caDNAno and Visual DSD, two DNA self-assembly models were described, which are respectively corresponding to Zhengzhou University of Light Industry badge and the prime number judging. After that, the application and feasibility of related softwares was discussed. And then, it was pionted that software platforms design was an important issue in the research of DNA selfassembly computation.

Key words: DNA computing; DNA self-assembly; simulation software

0 引言

自组装 DNA 计算是一种以分子生物学理论为 月

基础、以 DNA 分子为材质、以先进分子生物技术为 依托的新颖计算范式. 1994 年, L. M. Adleman^[1]利 用现代分子生物技术,成功解决了含有 7 个顶点的

收稿日期:2013-11-15

基金项目:国家自然科学基金项目(61070238,61272022,U1304620);河南省基础与前沿技术研究计划项目(122300413211,132300410183);河南省创新型科技人才队伍建设工程支持项目(124200510017);郑州市科技人才队伍建设计划(科技领军人才)项目(131PLJRC648);河南省教育厅科学技术研究重点项目(13A413371);郑州轻工业学院博士基金项目(BSJJ2010002)

作者简介:王子成(1976—),男,河南省永城县人,郑州轻工业学院讲师,博士,主要研究方向为分子电子学、智能计算. 通信作者:崔光照(1957—),男,河南省洛宁县人,郑州轻工业学院教授,博士,主要研究方向为生物计算与信息安全. 有向哈密尔顿路径问题,开启了 DNA 计算的新 纪元.

DNA 分子具有特殊的空间构型及物化特性. 分子内碱基组合方式的多样性使其具备了海量信息存储能力,分子间 Waston-Crick 碱基匹配机制使其具有信息读取与传递时的高度并行处理能力. 同时,DNA 分子间的识别过程还具有低能耗和强抗电磁干扰等优点. 自组装 DNA 计算模式的研究涵盖了生物学、信息学及纳米科学等多个学科的内容, 使其成为一个多学科交叉研究的前沿课题.

20世纪末至今,自组装 DNA 计算在理论研究 及技术实现方面都取得了长足进步.当前,自组装 DNA 计算研究主要有3种类型,分别是 DNA Tile 自 组装、DNA 折纸术及 DNA 链置换技术,其中链置换 技术是在自组装机理基础上最新出现的一种动态 自组装模式.

DNA 自组装体系中,在非共价键作用下,作为 基本结构单元的 DNA 分子能够自发组织或聚集成 具有稳定规则几何外形的有序功能结构.根据这一 特征,研究者开展了具有特定 DNA 纳米结构体系的 构筑研究. T. J. Fu 等^[2]认为以十字叉状的 Holliday 为中间体,通过具有黏性末端 DNA 分子间的结合可 实现更高维度的纳米结构的组装. C. Mao^[3]将其用 于 DNA Tile 自组装技术,开发出结构稳定的 DX Tile,并构建出多种计算模型. 2000 年, C. Mao 等^[4] 使用 Tile 自组装技术设计出求解累积异或的运算模 型,并通过实验证实了其可行性. 2002年, A. Carbone 等^[5]利用 DNA 分子自组装构建了布尔逻辑电 路模型. 2006 年, Y. Brun^[6]提出了基于 Tile 自组装 的加法和乘法运算模型. 2009 年, C. X. Zhang 等^[7] 采用 Tile 自组装技术设计了除法和减法运算模型. 随后,多种二维及三维组装结构模型相继出 现^[8-10],其中 DNA Tile 自组装因其成本低、操作简 便而得到广泛应用.

DNA 链置换技术是分子计算领域中新兴的技术,是指在部分杂交的 DNA 分子中,利用 DNA 分子间不同的结合力、分子体系的熵变及其能量起伏等因素,在不借助任何外力和催化剂的情况下,通过输入一条 DNA 单链而释放另一条 DNA 单链的过程.该技术可实现较长的双链杂交对较短杂交区域的取代,被置换出的短链可作为逻辑计算的输出信号.

2006年,S. Georg 等^[11]基于链置换技术设计出 与门、或门、非门等逻辑模块,并组合成逻辑电路. 2010年,L. L. Qian 等^[12]改进单个门单元设计,使逻 辑门设计更加模块化和标准化,为组建大规模电路 做了铺垫.2011年,L.L.Qian等^[13-14]采用72种 130条DNA链构建了具有计算4位二进制平方根 能力的逻辑电路;利用112种不同的DNA链组成了 4个相互联系的人工神经元网络,通过猜心术游戏 证明用DNA分子组成的神经网络具有一定的逻辑 推理能力.

DNA 分子自组装体系中,分子间进行信息获取 与传输时,依赖微观领域的能量起伏及微弱键合作 用. 自组装 DNA 计算研究以生化试验手段作为计算 过程的实现载体,需要一定的科研投入. 开发用于 分子组装体系仿真的软件平台能够降低科研成本, 提高工作效率,实现对自组装体系的可行性分析, 同时还可借助仿真结果预测模型结构,提升实验的 成功率.

本文将针对 caDNAno 工具、Visual DSD 工具、 NUPACK 3 种通用性较强的软件平台,综述各个软 件工具的主要功能、操作方法、使用范围以及笔者 所在课题组使用这些软件开展的实验验证等工作.

1 自组装 DNA 计算仿真软件

1.1 caDNAno 工具

caDNAno 是一个用于 DNA 分子绘图的软件,适 于构建二维及三维自组装 DNA 结构模型. 最早的 caDNAno 版本由哈佛大学 Wyss 生物工程研究所生 理学家 Shawn Douglas 开发.

caDNAno 工具的交互界面由 3 个可视窗口以及 各窗口周围的功能选项组成.窗口 I 中可显示分子 模型的侧面,用于确定模型的维数结构.它有 Honeycomb 和 Square 2 种模式:Honeycomb 用于构建蜂 窝状及管道状等结构;Square 用于构建二维平面及 三维立体结构等.窗口 II 中可显示 DNA 分子的长 度、序列及分子间的结合方式等信息,通过切换 Select,Pencil,Break 等功能键可完成绘图过程.选择 Pencil 键可绘制表示 DNA 分子的彩色直线;利用 Select 键移动或删除 DNA 分子,可改变分子长度; 通过 Break 键可实现 DNA 分子链的切断功能.窗口 III 为 caDNAno2 新增的 3D 视觉窗口,借助窗口中的 视角转换器可从不同视角观察 DNA 分子模型,形象 直观.

caDNAno软件平台下,笔者所在课题组以郑州 轻工业学院(以下简称轻院)校徽为对象,设计其二 维 DNA 自组装结构模型.该设计将1条含7000多 个碱基的 M13mp18 DNA 单链和 200 多条与之不同 部位互补的短 DNA 单链进行组装,生成直径约为 140 nm 的表征轻院校徽的 DNA 分子结构.其中 M13mp18 单链为脚手架链,短的 DNA 单链为订书 钉链.通过将脚手架链在水平方向上反复折叠形成 轻院校徽的形状,订书钉链在适当的位置对脚手架 链进行固定.

该设计过程包括以下2个步骤.

1)设计脚手架链的折线图.因二维图形的侧面 为直线,在软件的窗口 I 中 Square 模式下,选取任 一直线来确定二维结构.双链 DNA 的一个螺距约为 3.6 nm,10.5 个碱基,宽度约为2 nm,平行的 DNA 双链的间距约为1 nm.在等比例设计纳米级图形时 考虑到这些数据,同时还要考虑脚手架链的长度为 7 000 多个碱基,由此估算出设计的图形直径约为 140 nm.设计的模型高度约为47 nm×3 nm,可看 出,模型高度约为47 层 DNA 螺旋.将脚手架链折叠 成轻院校徽的形状时,为保证 DNA 链在折叠处的扭 力最小,每一段 DNA 双链的长度必须是半个螺距的 整数倍^[8].校徽形状的脚手架链折线图如图 1 所 示.考虑到校徽是镂空结构,为提高纳米结构的稳 定性,在原图形基础上增加3个连接处,这样既增强 结构稳定性,也不太影响图形美观.

2)设计订书钉链序列.每条订书钉链通过与多 层 DNA 螺旋相结合来实现对脚手架链的固定.已知 脚手架链的序列,根据 Watson-Crick 原则完成订书 钉链的序列编程.绝大多数订书钉链设计如图 2 所 示 7 种情况(箭头所指方向为5'→3').为保证订书 钉链有效地固定图形结构,除了边缘处,每个订书 钉链均连接 3 层 DNA 螺旋,如图 2 中的 a)和 b)为 图形中间使用订书钉链,其链长为 32 个碱基,其他 5 种为用于边缘部分的订书钉链.在设计订书钉链 时仍要保证 DNA 链折叠处的扭力最小.

在 caDNAno 软件平台中设计的郑州轻工业学院校徽纳米结构如图 3a)所示,3D 视觉窗口可显示校徽的 DNA 分子的组装仿真效果.图 3b)是图 3a) 中①区域放大图,可以看到选择的高度为 47 个双链 DNA 宽度.图 3c)为图 3a)中②区域放大图,其中蓝 线表示脚手架链,其他颜色直线为订书钉链.订书 钉链与脚手架链形成互补结构,完全填充图形结构.该设计构造出非对称的复杂的二维形状,体现 DNA 自组装具有构造几乎任何复杂二维纳米结构 的能力,为自组装 DNA 计算提供了有力的理论 依据.

1.2 Visual DSD 工具

由 L. Matthew 等^[15]设计的 Visual DSD 工具可用于构建 DNA 链置换计算模型,分析模型可行性.



图1 郑州轻工业学院校徽的脚手架链折线图





Visual DSD 以 Silverlight 交互式平台为编写基础,可 在 Windows, Mac OS X 和 Linux 操作系统下运行. 软 件由设置区、编码区和显示区组成. 在设置区对计 算模型进行仿真时,不同语法模式和仿真模式可得 到不同的仿真结果. 模型设计过程中通过编译可完 成模型修正,而后通过仿真可得组装模型仿真结 果. 在编码区完成 DNA 分子结构设计,设定其物质 的量、反应时间、数据采集次数及分子结合和分离 时间等参数,同时可实现分子子序列设计,也可采 用 Visual DSD 工具所提供的序列. 显示区保存了系 统编译、仿真和分析后的数据,能够显示仿真过程 中链置换过程、DNA 序列物质的量的变化曲线、反 应过程的始态和终止态,同时给出了相应的 Matlab 程序源代码,可借助 Matlab 仿真环境,对两者的仿 真结果进行比较,证实 Visual DSD 工具的可行性.

笔者所在课题组基于 DNA 链置换级联反应机 理,设计了用于素数判断的 DNA 分子计算模型,该 模型可以判断 4 位二进制数是否是素数,并采用 Visual DSD 工具进行置换过程的仿真.

素数是指仅能被1和其自身整除的自然数,对 素数判定的研究有重要意义.密码学认为,信息编 码时引入素数可实现信息的有效加密;生物界多数 生物的生命周期为素数(单位为 a),故可以最大程 度地避免碰见天敌等.本课题组设计的素数判断逻 辑电路如图4所示.该逻辑电路包括了与、或、非3种





图 3 caDNAno 软件平台中 郑州轻工业学院校徽设计图形



图4 用于素数判断的逻辑电路

基础逻辑门,可对4位二进制数进行素性判断. *I*₀— *I*₃ 为输入信号,*p* 为输出信号. 若输入为素数时,输 出高电平,*p*=1;否则,输出低电平,*p*=0. 将该逻辑 电路转换成分子电路,可实现素数判断的功能. 若 输入为素数时,可生成 DNA 链;否则,无 DNA 链 生成. 采用 Visual DSD 工具对该过程进行了仿真.应 用 Visual DSD 软件仿真时,需要设定模拟环境,设 置语法模式及仿真模式.在默认模式及随机模式 下,通过编码完成链置换分子模型的设计,部分程 序代码如下:

directive duration 20000.0 points 1000 directive plot $\langle p \rangle$; $\langle t^{\uparrow} \rangle$; $\langle t^{\uparrow} \rangle$; $\langle t^{\uparrow} \rangle$; directive scale 100.0 def bind = 0.0003 (* /nM/s *) def unbind = 0.1126 (* /s *) new t@ bind, unbind

上述代码中仿真时间为 20 000 s,数据采集次 数为1 000 次,物质的量变化监测对象分别是输出 链 < *p* > 和输入链 < *t*[^] _ > 、< *t*[^] _ *t*[^] > ,物质的量的 刻度标尺以及 DNA 结合和分离速率. 仿真结果如图 5 所示.

图 5 中共有 15 种情况,分别对应 1—15 之间的 自然数.其中 N 表示输入,p 表示输出状态,p = 0 表 示输出为低电平,即输入为非素数,p = 1 表示输入 为素数.红线条表示 p 物质的量,当输入为 2,3,5, 7,11,13 时,p 物质的量随时间不断增加,最终趋于 稳定,稳定后物质的量接近 1 000,而其他输入情况 p 物质的量保持不变.可见,在自然数 1—15 中,数 2,3,5,7,11,13 是素数.

该仿真实验中,链置换计算模型由 25 条 DNA 单链组成,结构复杂,使用 Visual DSD 工具大大降 低了工作量,提高了工作效率,最难得的是可以通 过仿真结果来判断模型设计的优劣,从而完善并优 化设计.

1.3 NUPACK 工具

NUPACK 由加州理工学院 Pierce 实验室开发, 是一个用于 DNA 序列分析和设计的软件,该工具是 免费和开源的.

NUPACK 工具初始界面包括 Analysis, Design, Utilitiesp 这 3 个选项. 基于分子热动力学理论, 可在 Analysis 中仿真不同温度、不同浓度下 DNA 链或 RNA 链的杂交反应, 仿真结果为热力学平衡态; 对 于多条 DNA 链或 RNA 链杂交后的结构, Design 中 将给出对应于该结构的最优分子编码序列; Utilities 选项融合了上述 2 项功能, 增加了更多功能设置, 可 仿真分子的不同视觉效果, 同时可对组成 DNA 序列 的碱基进行编码.

基于 DNA 链置换反应的与逻辑门反应过程,如 图 6 所示. 首先在与逻辑门中加入输入信号链 <1 2 >



图 5 用于素数判断的分子逻辑模型 Visual DSD 仿真

和链 <3 4 >,链 <1 2 > 和链 <3 4 > 和与逻辑门进 行链置换反应,最终替代并产生输出信号链 <2 3 >,实现与逻辑运算.

借助 NUPACK 工具的 Design 功能,可以完成对 图 6 所示与逻辑门的 DNA 链编码序列. NUPACK 工

具给出与逻辑门结构对应的序列见表 1, Visual DSD 平台中提供的与逻辑门的各个区域 DNA 序列见 表 2.

在 25 ℃条件下,上述 2 种序列的与逻辑门在 NUPACK 工具中的仿真结果如图 7 所示. NUPACK

工具设计序列的与门和2条输入链反应的仿真结果 如图7a)所示,输出信号链<23>的产率为47%. Visual DSD工具提供序列的与门在NUPACK平台 上和2条输入链反应,仿真结果如图7b)所示,输出 信号链<23>的产率为92%.



图6 与逻辑门反应的基本过程

表1 采用 N	UPACK 工具设计的 DNA 序列
DNA 链	序列(5′→3′)
<23>	CGATACCCATCGGTCTCTGGCACCCC TCCATTGCAGGACG
<1* 2* 3* 4* >	CCATCGTCCTGCAATGGAGGGGTGC CAGAGACCGATGGGTATCGATCGGA
表2 Visual DSD	平台中用于与门组装的 DNA 序列
序号	序列(5′→3′)
1	TATTCC

4	GCTA
2	CCCAAAACAAAACAAAACAA
3	CCCTTTTCTAAACTAAACAA

仿真结果表明,相对于 Visual DSD 仿真平台, 用 NUPACK 提供的序列能得到更加稳定的结构,并 且采用 Visual DSD 设计序列时,为防止 DNA 链自身 生成次级结构,减少 DNA 链间的错配率,需考虑的 因素较多,如每条 DNA 链仅用 3 种碱基,不同链的 5 和 3 端都设计成相同序列等;而用 NUPACK 设计 序列时,仅给出了 2 条链杂交时的最优情况.

此外,结构稳定的逻辑门不一定适合链置换反应. 链置换反应是一个不同链之间替代的过程,对链置换反应过程中的 DNA 序列进行编码时,不仅要考虑逻辑门结构的稳定性,还要考虑替代产物的稳定性,减少替代过程的可逆反应. Visual DSD 工具中提供序列的与门,在 Visual DSD 和 NUPACK 工具中仿真的输出信号产率同为 92%,2 个工具可互相证

strand3	0.47	μM		-	2	3	
	0.47	МЧ		1*	2*	3*	4*
strand1-strand2-strand4				1	2	3	4
	0	. 52	μМ	1*	2*	3*	4*
strand3-strand4					2	3	
strand1	0	. 52	μM	-	1	2	
		0. 53	μM	-	3	4	

a)表1中序列与门2输入NUPACK仿真结果

strand1-strand2-strand4			1	2	3	4
	0.92	μШ	1*	2*	3*	4*
strand3	0.92	н 11	_	2	3	
strand2				3	4	
strandi		-		1	2	
■ 0.073 µM strand3-strand4		_		2	3	
0.073 μM			1*	2*	3*	4*
strand1-strand3-strand4			1	2	3	
0.004 µM			1*	2*	3*	4*

图7 2种序列与逻辑门在 NUPACK 的仿真结果

明其可靠性.

可见,采用 NUPACK 平台进行仿真时,可设计 表征 DNA 分子结构的最优序列,并给出可能的结 构,还可得到不同链之间的结合率,是研究自组装 DNA 计算过程有力且可靠的仿真工具.

2 结论

自组装 DNA 计算研究以生化试验为基础,需要 一定的科研投入为前提.同时,作为纳米级微观世 界的探索性研究课题,有关自组装 DNA 计算模型的 可行性与正确的组装结果间存在较大的差距,同时 DNA 分子间的组装结果受控于诸多因素.所以,开 展生化试验前,借助相关的软件平台,对自组装 DNA 计算模型的可行性及组装结果的准确性进行 仿真,有利于开展模型的可行性分析及分子体系组 装结果的预测,有助于在试验前及时修正组装模 型,改进序列设计算法.通过这种方式,可以降低实 验室耗材使用,提高科研工作效率.但在使用过程 中,这些软件工具也存在一定的缺陷和局限性:采 用 caDNAno 软件进行模型结构设计时,只有 Honeycomb 和 Square 这 2 种方式,无法仿真多数简单模 型;在 Visual DSD 软件环境中,对带有发卡结构的 DNA 分子的组装过程却无法仿真; NUPACK 工具无法给出具有交叉结构的分子组装模型, 而且, 当分析的分子链的数目在5条以上时, 分子结合的仿真过程需要很长时间.

除上述软件平台外,有关学者也开发了其他用 于自组装 DNA 计算分子系统仿真的工具.如 SARSE 软件主要用于 DNA 折纸术研究,可直接给 出二位维图形的 DNA 折纸结果,也可开展三维模型 的设计研究;DomainDesign (DD)软件用于 DNA 链 置换级联反应,能够弥补 Visual DSD 不能设计发卡 结构的不足;SeesawCompiler 软件在 Visual DSD 工 具上设计翘板电路时更为快捷;KineticDesign (KinD)是一个用于序列设计和分析的工具,与 NU-PACK 相比,其计算速度较慢.

随着自组装 DNA 计算研究的逐步深入,上述软件的不足和局限性逐渐凸显出来,需要在实践中努力改进和完善.而且,随着 DNA 计算的蓬勃发展,会有许多新的生物技术和研究方法出现,这就要求开发更为高效、实用的新软件.

参考文献:

- [1] Adleman L M. Molecular computation of solutions to combinatorial problems[J]. Science, 1994, 266(5187):1021.
- [2] Fu T J, Seeman N C. DNA double-crossover molecules[J]. Biochemistry, 1993, 32(13):3211.
- [3] Mao C, Sun W, Seeman N C. Designed two dimensional DNA holliday junction arrays visualized by atomic force microscopy[J]. Journal of the American Chemical Society, 1999, 121(23):5437.
- [4] Mao C, LaBean T H, Reif J H, et al. Logical computation using algorithmic self-assembly of DNA triple-crossover molecules [J]. Nature, 2000, 407(6803):493.

- [5] Carbone A, Seeman N C. Circuits and programmable selfassembling DNA structures [J]. Proceeding of the National Academy of Sciences of the United States of Anerica, 2002,99(20):12577.
- [6] Brun Y. Arithmetic computation in the tile assembly model, addition and multiplication [J]. Theoretical Computer Science, 2007, 378:17.
- Zhang X C, Wang Y F, Chen Z H, et al. Arithmetic computation using self-assembly of DNA tiles: Subtraction and division [J]. Progress in Natural Science, 2009, 19 (3):377.
- [8] Rothemund P W. Folding DNA to create nanoscale shapes and patterns[J]. Nature, 2006, 440 (7082):297.
- [9] Dongran H, Suchetan P, Jeanette N, et al. DNA origami with complex curvatures in three-dimensional space [J]. Science, 2011, 332:342.
- [10] Bryan W, Mingjie D, Peng Y. Complex shapes self-assembled from single-stranded DNA tiles [J]. Nature, 2012, 485;623.
- [11] Georg S, David S, David Y Z, et al. Enzyme-free nucleic acid logic circuits[J]. Science, 2006, 314:1585.
- [12] Qian L L, Winfree E. A simple DNA gate motif for synthesizing large-scale circuits [J]. J R Soc Interface, 2010, 5347:70.
- [13] Qian L L, Winfree E. Scaling up digital circuit computation with DNA strand displacement cascades [J]. Science, 2011, 332(6034):1196.
- [14] Qian L L, Winfree E, Bruck J. Neural network computation with DNA strand displacement cascades [J]. Nature, 2011,475:368.
- [15] Matthew L, Simon Y, Filippo P, et al. Visual DSD: A design and analysis tool for DNA strand displacement systems[J]. Bioinformatics, 2011, 27(22):3211.