第 28 卷 第 10 期
2005 年 10 月
计 算 机 学 报
CHINESE JOURNAL OF COMPUTERS
Vol. 28 No. 10
Oct. 2005

DNA 计算机:原理、进展及难点(Ⅱ)
计算机"数据库"的形成——DNA 分子的合成问题

许 进 黄布毅
(华中科技大学分子生物计算机研究所 武汉 430074)

摘 要 基于生化反应机理的 DNA 计算机模型引起了科学领域内许多不同学科学者们的关注与兴趣. DNA 计算已经成为国际科学研究前沿领域内的一个新热点. DNA 计算机的研制需要诸如生物工程、计算机科学、数学、物理、化学、信息科学、微电子技术、激光技术以及控制科学等许多学科的共同协作攻关. 作者以系列文章的形式拟对 DNA 计算机的基本原理、研究进展、DNA 计算的模型以及当前研究中的难点给予研讨. 该文属第二篇,重点讨论 DNA 计算机研制中 DNA 分子的合成问题. DNA 分子的合成问题不仅是 DNA 计算中生物操作过程首先要处理的问题,而且是 DNA 计算机研制中必须要解决的问题,因为最终实用化的 DNA 计算机应是一种全自动化的. 如何将 DNA 分子的合成过程与编码、其它生化操作自动地衔接起来是全自动化 DNA 计算机当前研究的关键难题. 若要解决这个问题,人们必须很熟悉有关 DNA 分子合成的基本原理以及合成技术. 这也是该文的动机.

关键词 DNA 计算;DNA 分子的合成;化学合成;基因合成;POA 技术
中图法分类号 TP301

DNA Computer Principle, Advances and Difficulties (Ⅱ): Setting up the Database of DNA Computer by the Synthesis of DNA Molecules

XU Jin HUANG Bu-Yi
(*Institute of Biomolecular Computer, Huazhong University of Science and Technology, Wuhan* 430074)

Abstract Biomolecular computing is computation at the molecular scale, using biotechnology engineering techniques. Recently, many scientists in different fields are interest in DNA computer model based on reaction of biochemistry because DNA computer has formed a new science field. The studying and making DNA computer needs many science subjects such as biological engineering, computer science, mathematics, physics, chemistry, information science, microelectronics, laser technology, and control science, etc. Based on this, the advances of DNA computer are considered in detail in this four-part paper such as fundamental principle, several models of DNA computing, and difficulties. This paper is the second. In this paper, the authors discuss the synthesizing methods of DNA molecules in DNA computing. The problem of synthesizing DNA molecule not only is the first problem needed to be processed in biological operations of DNA computing, but also must be solved problem in making DNA computer because the final DNA computer should be automatic. Recently, The key problem studied in DNA computer is how we make automatic DNA computer based on encoding, synthesizing DNA molecules, biological operations. For this, we must be familiar with the basic principle and technology in synthesizing DNA molecules, this is the motive of the paper.

Keywords DNA computing; synthesis of DNA molecules; chemic synthesis; gene synthesis; POA(Parallel Overlap Assembly) technology

收稿日期:2005-06-14;**修改稿收到日期**:2005-08-09. **本课题得到国家自然科学基金**(60274026,30370356,60373089,60403002)**资助**.
许 进,男,1959 年生,教授,博士生导师,华中科技大学分子生物计算机研究所所长,主要研究领域为 DNA 计算机、图论、神经网络、遗传算法等. E-mail:jxu@mail.hust.edu.cn. **黄布毅**,女,1959 年生,博士,教授,研究领域为 DNA 计算机、控制科学等.

1 引 言

DNA 计算必须有生物操作,生物操作可以没有生物酶,但必须有作为信息处理"数据"的 DNA 分子. 在 DNA 计算中,作为"数据"的 DNA 分子不能随机地产生,原因是由于 DNA 分子在杂交过程中可能产生不希望出现的假阳性和假阴性 DNA 分子. 为了避免不必要的分子杂交现象,需要进行优化编码. 在编码过程中,存在问题的规模与 DNA 分子长度的选择问题(如何给出最优长度的 DNA 序列是一个组合最优化问题)等. 在确定了问题的编码而需进行生化操作时,首要的问题是对这些编码的 DNA 链进行合成,如何合成?目前主要是在有些生物制品公司用合成仪进行合成,所用的合成仪有 ABI 公司生产的老式的 392 和 394 型合成仪[1],也有该公司生产的较先进的 3900 合成仪. 而有些学者则是利用一些新的研究成果进行合成,如 Ouyang 等人[2]在建立求解图的最大图问题的 DNA 计算模型时采用了由 Stemmer 等人[3~5]建立的平行重叠组装(Parallel Overlap Assembly,POA)技术进行. 当然,我们期望的是采用"电子计算机"方法对 DNA 分子来进行合成:"对编了码的 DNA 序列,按其顺序,只需要单击电子计算机键盘上的 4 个英文字母 A,C,G,T 即可合成". 这种设想目前还不能实现. 何时能实现?用什么方法来实现?将 IT 技术与 DNA 计算机有机结合起来,是使 DNA 计算机走向"成熟"的关键.

将 IT 技术,特别是电子计算机应用于 DNA 计算机,使得 DNA 计算机走向全自动的实用化似乎是一条"必由之路",也是一条"艰难之路". 要完成这个目标,我们不仅要非常熟悉有关 DNA 分子合成的基本原理与技术,而且要对相关的 IT 技术中有关控制系统、电子计算机的基本原理熟练掌握. 本文将主要对前者,即 DNA 分子合成的有关基本原理与技术给予较为详细的讨论.

DNA 分子的化学合成的开拓者是 Khorana,他在 1967 年就提出了用人工合成基因,即化学方法合成基因的想法,并赋予实践[6]. 1979 年,Khorana 在《Science》杂志上发表了题为"一个基因的合成"的著名论文,为整个基因工程的发展起到了奠基的作用. 也促进了 DNA 计算机的研究. DNA 计算机所用的"数据"是大量的 DNA 分子. 各种各样的 DNA 分子的合成,只有当今的自动化合成仪才能"胜任"此项庞大的工作.

DNA 分子合成主要有 5 种方法:磷酸二酯键方法,磷酸三酯键、亚磷酸三酯键方法以及在后二者基础上发展起来的固相合成法和自动化合成法等.

2 核酸的生物合成

DNA 计算所需要的 DNA 分子的合成是化学合成方法,化学合成是在单个核酸的基础上进行的. 因而了解核酸的生物合成过程对于未来全自动化 DNA 计算机的研究是有意义的. 故在本节里,我们将简要地介绍核酸的生物合成. 当然,在自然界,动物、植物和微生物体中通常都能本能地合成各种嘌呤和嘧啶核苷酸. 其生物机理本节将给予简要介绍.

DNA 计算中所用的"数据"是 DNA 分子,因而本文主要关注有关 DNA 分子的合成. DNA 分子的基本单位是脱氧核糖核苷酸,而在生物体内脱氧核糖核苷酸的合成是通过核糖核苷酸还原而成,其中腺嘌呤、鸟嘌呤和胞嘧啶脱氧核糖核苷酸经过还原后,将核糖中第二位 C 上的氧脱去,形成相应的脱氧核糖核苷酸;胸腺嘧啶脱氧核糖核苷酸的形成需要两步:(1)由尿嘧啶核糖核苷酸还原成尿嘧啶脱氧核糖核苷酸;(2)将尿嘧啶甲基化转变成胸腺嘧啶. 下面,我们仅以嘌呤核苷酸的合成为例,介绍核糖核苷酸的生物合成,而嘧啶核苷酸的合成略去.

嘌呤核苷酸的合成

在生物体内,采用二氧化碳、甲酸盐、谷氨酰胺、天冬氨酸和甘氨酸等来合成嘌呤环. 嘌呤的分子结构如图 1(a)所示. 嘌呤环中的第 1 位 N 来自天冬氨酸的氨基;第 3 位及第 9 位 N 来自谷氨酰胺的酰胺基;第 2 位及第 8 位的 C 来自甲酸盐;第 6 位的 C 来自二氧化碳;第 4,5,7 位的 C 来自甘氨酸,如图 1(b)所示.

核糖核苷酸的生物合成过程是:从 5′-磷酸核糖焦磷酸开始,经过一系列的酶促反应,生成次黄嘌呤核苷酸,再转变成其它的嘌呤核苷酸. 在图 2 中,我们给出了磷酸核糖焦磷酸、次黄嘌呤核苷酸的分子结构. 在图 3 中,我们给出了嘌呤核苷酸合成的主要过程. 其中次黄嘌呤核苷酸的合成过程需要 10 步化学反应,限于篇幅,我们在这里不作详述,有兴趣的读者可参阅文献[8].

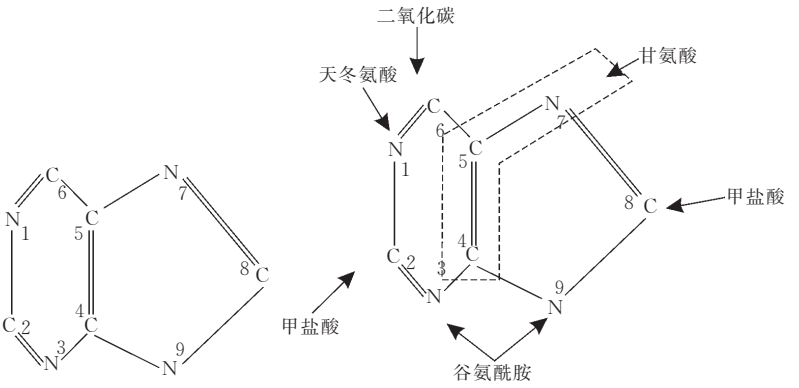

图 1　嘌呤环以及合成它的元素来源

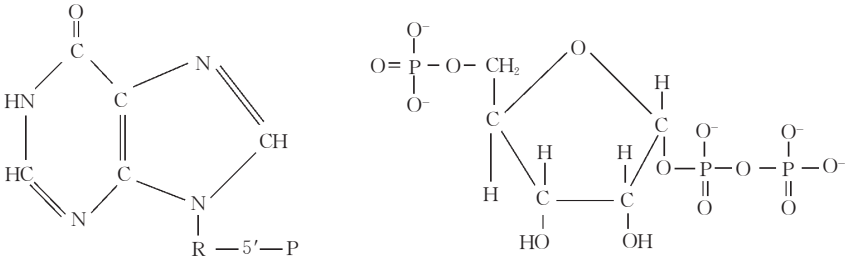

图 2　次黄嘌呤核苷酸、5′-磷酸核糖焦磷酸的分子结构

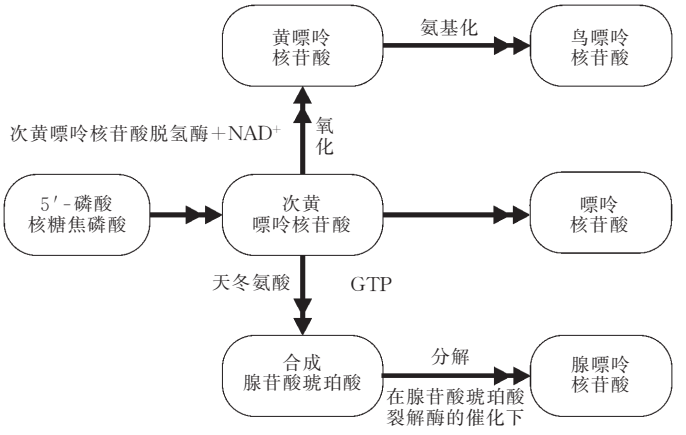

图 3　嘌呤核苷酸合成的主要过程

在次黄嘌呤核苷酸的基础上，生物体内由次黄嘌呤核苷酸氨基化生成腺嘌呤核苷酸. 其步骤如下：

1. 次黄嘌呤核苷酸在 GTP 供给能量的条件下与天冬氨酸合成腺苷琥珀酸；

2. 腺苷琥珀酸在腺苷琥珀酸裂解酶的催化下分解成腺嘌呤核苷酸和延胡索酸等.

3　DNA 的化学合成

DNA 合成是指一个核苷上的活性 3′-磷酸基团与另一个 5′-的羟基偶联. 前者是在溶液中传送的单体，后者被固定在固相载体上，核苷酸间的键因此形成. 为了下一步偶联，使 DNA 链增长，必须有另外三个化学反应.

3.1　磷酸二酯法合成 DNA 分子

磷酸二酯法是由 Khorana 以及同事首先创立和发展起来的一种 DNA 化学合成方法. 磷酸二酯键法的基本原理是：将两个在 5′-或 3′-末端各带有适当保护基的脱氧单核苷酸连接起来，形成一个带有磷酸二酯键的脱氧二核苷酸分子. 合成所用的原料是：脱氧核苷酸或脱氧单核苷酸. 下面给出具体合成的方法和步骤.

1. 确定好要合成的 DNA 序列，并假定 4 种脱氧单核苷酸均有无限的数量. 对所有的 5′-和 3′-末端加保护（如图 4 和图 5 所示）；

2. 按照要合成的 DNA 序列，先对第一个要合成的脱氧

核苷酸脱去 3′-端的保护，对序列中的第二个则脱去 5′-端的保护，然后进行缩合反应（或称为偶联反应），形成二者之间的磷酸二酯键，所得的序列记为 X. 显然，序列 X 的 5′-端和 3′-端均受到保护；

3. 脱去 X 中的 3′-端的保护，并脱去要合成的下一个要合成的脱氧核苷酸的 5′-端的保护，进行缩合反应，仍将反应后的序列记为 X.

上述步骤 3 不断进行，直到要合成的 DNA 序列中的最后一个字母被合成后终止.

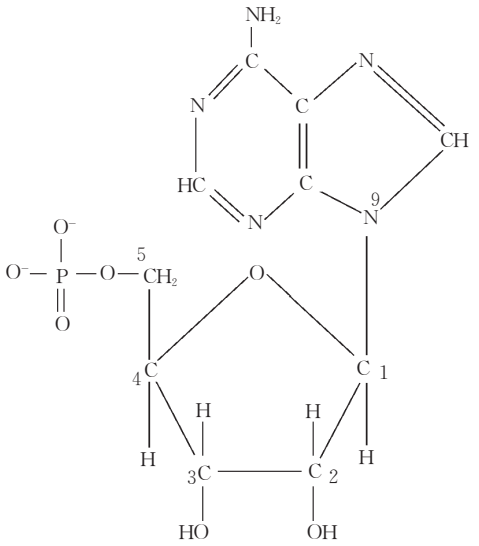

图 4 DMT（二甲氧基三苯甲基（4，4′-dimethoxytrityl））结构

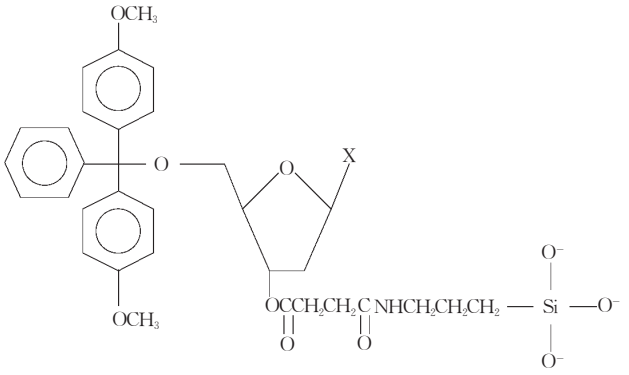

图 5 DMT-保护的核苷酸（其中 X 表示碱基 A，G，C 或 T[1]）

注 1. 保护基问题.

将不参加反应的集团用适当的保护基选择性地保护起来，其具体做法是：将保护基加到 5′或 3′的羟基端，通常采用的保护基有

（1）芳基磺酰氯（$ArSo_2CI$）；

（2）二环己己碳二亚胺（DCC）；

（3）二甲氧基三苯甲基（DMT）.

注 2. 磷酸二酯法合成 DNA 分子的不足：

（1）化学反应时间长；

（2）随着合成链的增长相应的产量逐渐下降，因而在最初用人工化学合成时，最多只能合成 15bp 的寡聚体；

（3）纯化步骤麻烦费时.

由于磷酸二酯法的上述缺点，它很快就被一些先进的方法所取代，这些方法将在下面给予简介.

3.2 磷酸三酯法合成 DNA 分子

磷酸三酯法与磷酸二酯法在方法与技术上几乎是完全一样的：它们均采用在单核苷酸的 5′-端和 3′-端加保护基团的方法；不同的是：在磷酸三酯法中，参加缩合反应的单核苷酸是一种核苷 3′-单磷酸的衍生物. 在采用磷酸三酯法原理合成 DNA 分子时，从技术上采用了一种称为固相合成的方法. 固相合成的原理是 20 世纪 50 年代由 Bruce Merrifiels 为了合成多肽链而创造的一种技术. 随后很快就将该方法应用于合成寡核苷酸[9]. 所谓固相合成法，是指将要合成 DNA 序列中的最后一个寡核苷酸固定在所谓的固相载体上，而把与它相连的，也就是进行缩合反应的单核苷酸通过溶液进行传送. 倒数第三个要被合成的核苷酸通过溶液与已经被合成在固相载体的倒数第二个进行缩合反应，如此反复进行，直到所有的 DNA 链被合成为止.

3.3 亚磷酸三酯（亚磷酰胺）法合成 DNA 分子

亚磷酰胺方法是目前寡核苷酸合成的主要方法，原因是其具有有效和快速的偶联以及开始物质的稳定性[10]. 合成的 DNA 链连接在载体上，使液相中过量的试剂过滤去掉. 因而在每个循环过程间不需要进行纯化. 其中的载体是一种硅胶形式的控制微孔玻璃珠.

亚磷酸三酯法合成寡核苷酸的每一个循环由脱 DMT、偶联、封闭和氧化等 4 个步骤构成. 这 4 个步骤不断重复，直到最后一个单核苷酸被合成. 当合成结束后，将所合成的 DNA 链从固相载体上切割下来并脱去保护. 这样所得到的目的 DNA 链就是 DNA 计算机存储库中所需要的. 进而，可根据不同的问题作进一步的生物操作.

下面，我们主要讨论基于亚磷酰胺法对 DNA 链进行化学合成的四个步骤. 这种合成是采用固相合成方法进行的.

1. 用酸处理衍生的固相载体以除去保护集团 DMT. 更具体地，用三氯乙酸（TCA）或二氯乙酸来处理，去掉 5′-端的保护基 DMT，露出 5′-OH，以便与相邻的单核苷酸进行偶联，图 6 给出此过程的一个说明. 合成的具体方法为：

1.1 用乙腈冲洗载体以除去上步留下的微量剂；

1.2 将 TCA 或二氯乙酸输送到合成柱子上，切下 DMT

集团，并用 TCL 冲洗多次，以便使 DMT 阳离子完全被冲洗下来，这是因为 DMT 阳离子具有高度的活性，可重新将任何活性的亲核试剂 DMT 化；

1.3 每次冲洗后均用氩气从柱的底部到顶部进行 DMT 冲洗；

1.4 将残留的 TCA 用乙腈洗涤.

2. 偶联反应. 偶联反应是用四唑这种弱碱作催化反应促使加进来的第二个单核苷酸同已经附着在固相载体上的核苷酸中暴露的 5′-OH 基进行缩合. 基本原理：亚磷酰胺与四唑被送到有载体键合的核苷酸柱内. 二异丙酰胺被质子化且被四唑取代，当 5′-OH 偶联到亚磷上去时，就形成了 5′→3′核苷酸间的键. 详细说明如图 7 所示. 其具体步骤如下：

2.1 用过量的乙腈对载体洗涤，对载体作无水处理，且作清除亲核试剂（如水）处理；

2.2 将合成柱子用氩气反冲干燥，除去残留的乙腈；

2.3 将亚磷酰胺活化剂四唑输送到柱子.

3. 封闭. 把所有没有参加偶联反应的 5′-OH 集团全部封闭起来. 其方法是加入乙酸酐.

4. 氧化反应. 由于在两个核苷酸之间新形成的 3′-5′亚磷酸三酯键不稳定，容易被酸和碱切断. 因此，封闭后的三价亚磷酸三酯立即被氧化成稳定的五价磷酸三酯.

值得注意的是，当亚磷没有氧化时不要终止合成. 当氧化结束后，核苷酸合成的一个循环随即结束. 寡聚物的 5′-端被 DMT 保护，再通过除掉 5′-端的 DMT，重复上述 4 个步骤，直到所需的 DNA 完全合成为止.

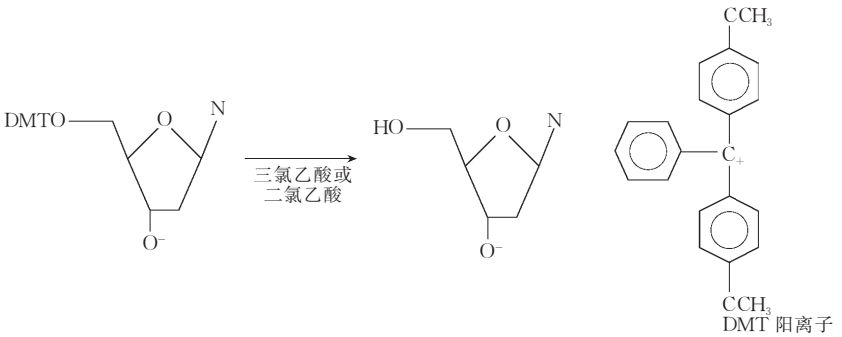

图 6　脱 DMT 示意图[1]

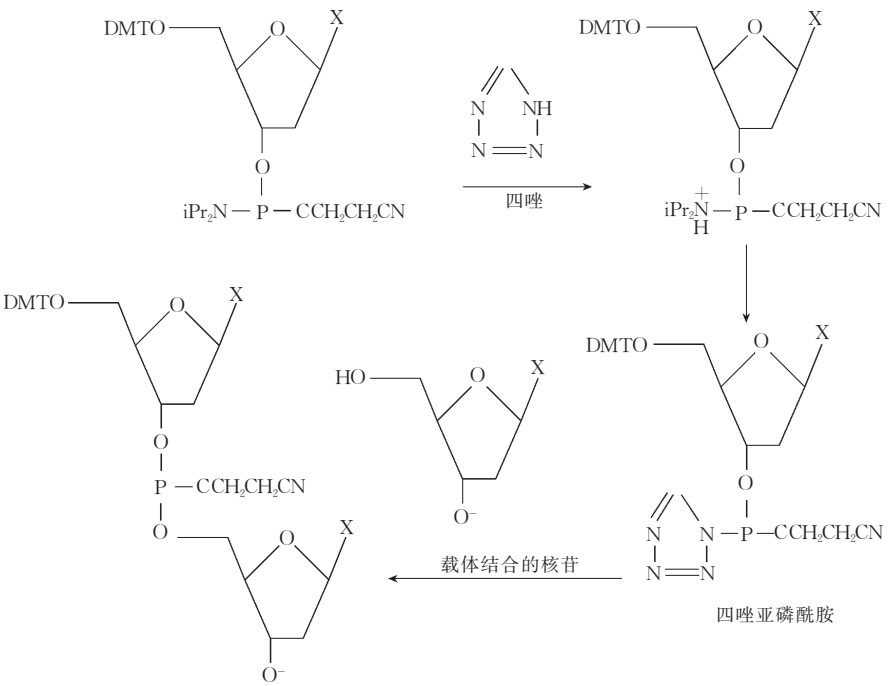

图 7　偶联基本原理图示[1]

注 1.　当寡核苷酸用凝胶电泳和离子交换 HPLC 纯化时，通常以 trityl-off 形式合成；而当用 OPC 或指定 DMT（trityl-specific）反相 HPLC 纯化时，其合成一般保留 DMT（trityl-on）.

注 2.　对于要合成的 DNA 序列，在合成之后一般可通过 PCR 扩增等方法来检测合成的准确性.

有关这方面的详细讨论可参见文献[1].

4 基因合成 DNA 方法

上述所讨论的化学合成方法合成 DNA 片断的能力在 150～200bp 之间. 然而在 DNA 计算机的研究中,一般都超过了这个范围,例如,在文献[11]中所建立的用于解决 20 个变量 3-SAT 问题的 104856 个长度为 300bp 的 DNA 序列;我们最近给出的用于求解具有 21 个顶点的图的 3-顶点着色问题的 10460353203 个长度为 315bp 的 DNA 序列①. 当然,随着 DNA 计算机研究的不断深入,所需要的 DNA 链会更长,并且量会更多,因而单纯用上述化学合成方法肯定是不行的. 而应该用所谓的基因合成寡核苷酸方法. 在这里,我们将这方面的主要研究概括如下基于寡核苷酸合成和制备 DNA 序列的方法,该方法主要是以连接酶为工具[12～17];1988 年发表的 FokI 方法[18];20 世纪 90 年代早期发表的自引物 PCR[19～22]方法;1995 年发表的 PCR 制备方法[23]和 2002 年的进一步改进[25];1996 年的模版定向配位连接(TDL)方法[24]. 最近又有两种合成和制备长 DNA 序列(约 5Kb)的方法问世. 一种是基于 PCR 的热力学内外平衡 TBIO 基因合成方法,它是利用新的引物设计方法实现长基因的高保真制备[26]. 另一种是 Smith 等人所提出的装备 PCR 方法的改进版本,它可以通过合成寡核苷酸来精确制备 5～6Kb 的 DNA 片段.

当前基因合成的主要途径是通过 DNA 合成仪来实现,关于这方面的理论方法可参见相关的文献,这里不再讨论.

5 POA 方法合成 DNA 分子

平行重叠组装方法 POA(Parallel Overlap Assembly)是一种合成 DNA 分子的方法,过去主要应用于诸如基因的构造、基因重组和 DNA 重组等[2]. 1997 年, Kaplan 等人将 POA 方法应用于 DNA 计算的研究之中[2,26]. 在这一节里,我们将给出 POA 方法合成 DNA 分子的主要原理,并通过实例给予讨论说明.

在合成过程中,先准备好一个初始有序的重叠低聚核苷酸(如图 8 所示),这样便于退火后低聚核苷酸可以通过一个 DNA 聚合酶得到扩增. 通过反复的融化、退火和扩增,就可以由小的片段构成长的 DNA 分子. 只有在 DNA 的每个片段都包含有它只

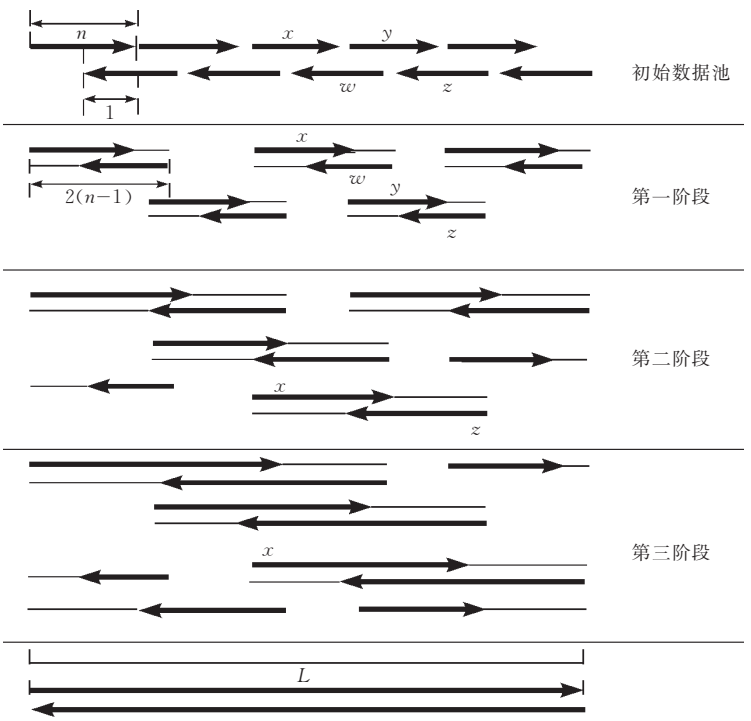

图 8 重叠扩增技术(一个长的 DNA 串,例如,从短片递归组装的一个基因,允许 3′端结束杂交,被聚合酶延伸. 粗箭头描述单链 DNA 进入反应的每一个阶段. 细箭头表示加入聚合酶的 DNA)

① Xu Jin, Qiang Xiao-Li, Fang Gang, Zhou Kang. A DNA computer for graph coloring. Submitted to some journal

能与邻近的片段退火的信息才能达到完全正确的组装，这样才能在最终的链中找到它的正确位置. 另一方面，如果一个片断与它的不邻近的片段退火，则 POA 会产生差异.

在 POA 中，通过最后分子长度的中间扩展进行并行扩增而得到长的分子（见图 8）. 随着 POA 的进行，每个分子伸长，变成最后产物的一个大的部分. 这种组合的动力学与 PCR 或者表示不同分析的不同的扩增方法相关但是不同. POA 的并行构建和 DNA 聚合酶只对多核苷酸的 3′端进行扩展的情况形成的一个直接结果是稀释和 PCR 的完全重叠扩增. 为了论证这一观点，考虑图 8 中标记了的分子 x. 通过组装的第 3 步，x 延伸到最终产物的 3′端. 但由于 5′端没有延伸，x 并不是我们期望的最终产物的成分. 另外，对图 8 的仔细检查可以看出随着链的增长，链的数目并没有发生变化. 在 10 个起始链中，只有从两个低聚核苷酸 5′端形成的两个链被延长以形成最终产物. 因此通过重叠扩增的组装必须通过稀释以移除不完全的产物和通过采用 5′终端的低聚核苷酸作为引物的 PCR 来充实完全分析的集合.

1997 年，Ouyang 等人在建立图的最大团问题的 DNA 计算模型时，对如图 9(a)所示的 6 个顶点的图进行了生化实验. 其中在数据池时采用了POA 技术. 文中所给出的算法共 4 步，其中的第一步是：对于 N 个顶点的图，每一个可能的团用一个 N 位二进制数来表示. 如果某一位的取值为 1，表示该顶点在所求的团中；如果一位的取值为 0，表示相应的顶点不在团中. 例如，团(4，1，0)被二进制数 010011 表示，而团(5，4，3，2)是图 9(a)中的最大团，由二进制数字 111100 表示. 这样，我们把 N 个顶点的图中所有可能的团的集合映射成了所有可能的 N 位二进制数的集合.

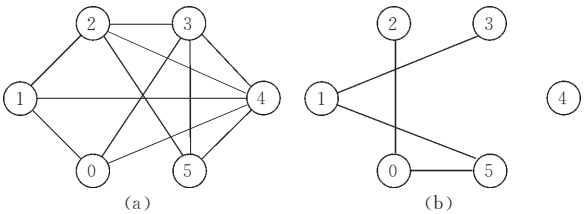

图 9　文献[2]中给出的实例图与它的补图

为了完成此步，采用 POA 技术，具体如下：以双链 DNA 的形式设计数据结构. 一个二进制数中的每一位数被两个 DNA 片段表示，一个叫做该位的值(V_i)，另一个叫做该位的位置(P_i). 对于代表 6 位二进制数的 DNA 分子来说，存在 6 个位值片段($V_0 \sim V_5$)及 7 个位置片段($P_0 \sim P_6$)，它们交替有序地排列(如图 10 所示).

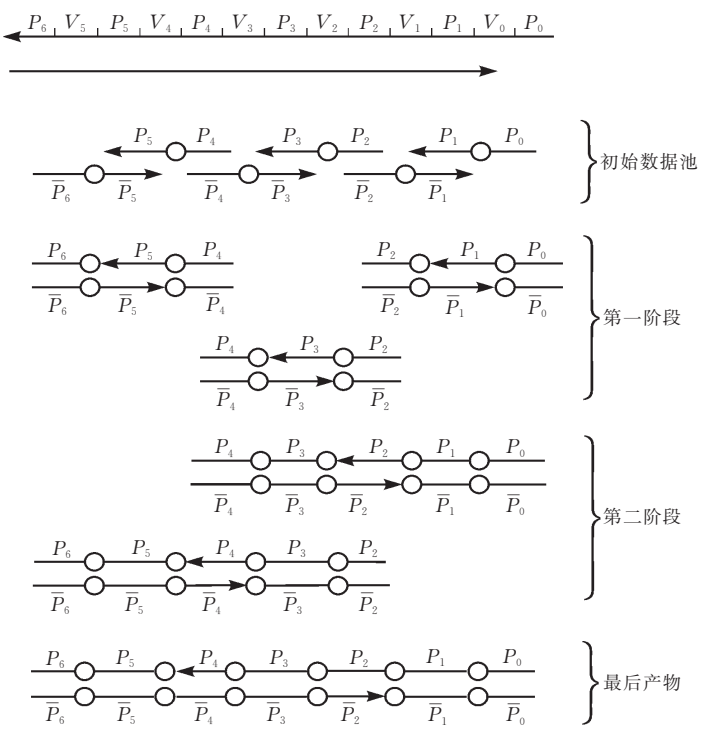

图 10　DNA 数据编码

最后的位置片段 P_6 对聚合链反应(PCR)放大是必需的. 如果 $V_i=1$，则 V_i 的长度为 0bp，如果 $V_i=0$，则 V_i 长度为 10bp，这样，最长的 DNA 链有 200bp，它对应数字 000000，而最短的 DNA 只有

140bp,对应数字 111111. 这种设计是简单而有效的,因为只需知道答案中数字 1 的个数就行了. 为了以后操作方便,我们在每个 $V_i=1$ 对应的序列中嵌上控制序列,以便利用相应的控制酶进行剪切. 每个低聚核苷酸由两个位置元和一个值元组成,当 i 为偶数时它为 $P_iV_iP_{i+1}$ 的形式,当 i 为奇数时它为 $\overline{P_{i+1}V_iP_i}$ 的形式,其中上面的横线代表补序列,V_i 的值可以是 0 或 1. 这 12 种片段核苷酸混合起来进行热循环,在每轮热循环中,一个核苷酸位置元上的链与另一个核苷酸上的补链退火结合,在聚合过程中,3′端逐渐延伸,形成一较长的双链 DNA. 经过几轮热循环之后,含有 $V_0V_1V_2V_3V_4V_5$ 的所有组合的数据池就构造出来了(见图 10).

6 总 结

本文主要讨论了 DNA 计算中作为"数据"的 DNA 分子的合成问题,这是每个 DNA 计算在实验或者生化操作中必须首先进行的问题. 文章从 DNA 分子合成的生化原理到 DNA 分子的化学合成原理,基因合成原理到 POA 合成都进行了论述. 由于当前的 DNA 计算研究主要是通过合成仪进行的,如 Adleman 等人关于 20 个变量 SAT 问题的求解以及我们最近建立的图的顶点着色问题的解空间的"数据"构造上,均采用了合成仪进行合成. 所以,在基因合成方面,主要给出了文献性的综述. 而对于 POA 技术,则通过 Kaplan 等人用于求解图的最大团问题进行了介绍.

希望本文的介绍能对 DNA 分子合成问题的进一步研究提供参考.

参 考 文 献

1 Edge Wang R., Xia Ling-Wei. The Synthesis, Marker and Applications of Nucleic Acid Probe. Beijing: Science Press, 1998(in Chinese)
(Edge Wang R.,夏令伟. 核酸探针的合成、标记及应用. 北京:科学出版社, 1998)

2 Ouyang Qi, Kaplan Peter D., Liu Shu-Mao, Libchaber Albert. DNA solution of the maximal clique problem. Science, 1997, 278(17): 446～449

3 Stemmer W. P. C.. DNA shuffling by random fragmentation and reassembly: In vitro recombination for molecular evolution. In: Proceedings of National Academy of Sciences, USA, 1994, 91: 10747～10751

4 Stemmer W. P. C.. Rapid evolution of a protein in vitro by DNA shuffling. Nature, 1994, 370(4): 389～391

5 Stemmer W. P., Crameri A., Ha K. D., Brennan T. M., Heyneker H. L.. Single-step assembly of a gene and entire plasmid from large numbers of oligonucleotides. Gene, 1995, 164(1): 49～53

6 Khorana H. G.. Total synthesis of a gene. Science, 1979, 203(4381): 614～625

7 Wu Nai-Hu. The Principle of Gene Engineering (Volume 1). Beijing: Science Press, 2001, 82～89(in Chinese)
(吴乃虎. 基因工程原理(上册). 北京:科学出版社, 2001, 82～89)

8 Shen Tong, Wang Jing-Yan. Biochemistry (Volume 2 of the Second Edition). Beijing: Higher Education Press, 2002, 302～315(in Chinese)
(沈 同,王镜岩. 生物化学(第二版下册). 北京:高等教育出版社, 2002, 302～315)

9 Letsinger R. L., Finnan J. L., Heavner G. A., Lunsford W. B.. Phosphite coupling procedure for generating internucleotide links. Journal of the American Chemical Society, 1975, 97: 3278～3279

10 Beaucage S. L., Caruthers M. H.. A synthesis of a gene based on phosphoramidite. Tetrahedron Letters, 1981, 22(19): 1859

11 Braich Ravinderjit S., Chelyapov Nickolas, Johnson Cliff, Rothemund Paul W. K., Adleman Leonard. Solution of a 20-variable 3-SAT problem on a DNA computer. Science, 2002, 296(19): 499～502

12 Smith J., Cook E., Fotheringham I. *et al*.. Chemical synthesis and cloning of a gene for human beta-urogastrone. Nucleic Acids Research, 1982, 10(15): 4467～4482

13 Edge M. D., Greene A. R., Heathcliffe G. R. *et al*.. Chemical synthesis of a human interferon-alpha 2 gene and its expression in Escherichia coli. Nucleic Acids Research, 1983, 11(18): 6419～6435

14 Jay E., MacKnight D., Lutze-Wallace C. *et al*.. Chemical synthesis of a biologically active gene for human immune interferon-gamma. Prospect for site-specific mutagenesis and structure-function studies. Journal of Biological Chemistry, 1984, 259(10): 6311～6317

15 Sproat B. S., Gait M. J.. Chemical synthesis of a gene for somatomedin C. Nucleic Acids Research, 1985, 13(8): 2959～2977

16 Ecker D. J., Khan M. I., Marsh J. *et al*.. Chemical synthesis and expression of a cassette adapted ubiquitin gene. Journal of Biological Chemistry, 1987, 262(8): 3524～3527

17 Ashman K., Matthews N., Frank R. W.. Chemical synthesis, expression and product assessment of a gene coding for biologically active human tumour necrosis factor alpha. Protein Engineering, 1989, 2: 387～391

18 Mandecki W., Bolling T. J.. FokI method of gene synthesis. Gene, 1988, 68: 101～107

19 Dillon P. J., Rosen C. A.. Arapid method for the construction of synthetic genes using the polymerase chain reaction. Bio-

techniques, 1990, 9: 298～300

20 Prodromou C., Pearl L. H.. Recursive PCR: A novel technique for total gene synthesis. Protein Engineering, 1992, 5(8): 827～829

21 Ciccarelli R. B., Gunyuzlu P., Huang J. *et al*.. Construction of synthetic genes using PCR after automated DNA synthesis of their entire top and bottom strands. Nucleic Acids Research, 1991, 19(21): 6007～6013

22 Hayashi N., Welschof M., Zewe M. *et al*.. Simultaneous mutagenesis of antibody CDRregions by overlap extension and PCR. Biotechniques, 1994, 17(2): 310～315

23 Strizhov N., Keller M., Mathur J. *et al*.. A synthetic cryIC gene, encoding a Bacillus thuringiensis delta-endotoxin, confers Spodoptera resistance in alfalfa and tobacco. In: Proceedings of National Academy of Sciences, USA, 1996, 93: 15012～15017

24 David M. H., Jacek L.. DNAWorks: An automated method for designing oligonucleotides for PCR-based gene synthesis. Nucleic Acids Research, 2002, 30(1): 43

25 Gao X., Yo P., Keith A. *et al*.. Thermodynamically balanced inside-out (TBIO) PCR-based gene synthesis: A novel method of primer design for highfidelity assembly of longer gene sequences. Nucleic Acids Research, 2003, 31(1): 143

26 Kaplan Peter D., Ouyang Qi *et al*.. Parallel overlap assembly for the construction of computational DNA libraries. Journal of Theoretical Biology, 1997, 188(7): 333～341

XU Jin, born in 1959, professor, Ph. D. supervisor. His research interests include DNA computing and DNA computer, neural networks, genetic algorithms, graph theory etc.

HUANG Bu-Yi, born in 1959, Ph. D., professor. Her interests include DNA computer, control science etc.

Background

This research is supported by the National Natural Science Foundation of China under grants (60373089, 30370356, 60274026): Research on the Theory, Model and Method of DNA Computer etc. The projects mainly focus on DNA computer models for processing graphical messages, including encoding DNA sequences, synthesizing DNA molecules, setting up the model, detecting solutions, etc. Our research group has been working on many aspects of DNA computing since 1996. We have published a monograph and more than 100 papers on DNA computing and DNA computer. In this paper, we will present a basic work of setting up the database of DNA computer, that is synthesizing DNA molecules.