# 基于DNA计算机求解二十个变量三可满足性问题

摘要：在一台简单的DNA计算机上以二十个变量三可满足性问题为例求解了NP完全问题。在对超过100万（220）种可能性进行详尽搜索（exhaustive search , 穷举搜索、暴力搜索）后，找到了唯一的答案。这个计算规模可能是迄今为止通过非电子手段解决的最大规模，并且似乎超出了没有辅助的人工计算的正常范围。

分子计算固有的巨大并行性、超高的能量效率和卓越的信息密度，提高了分子计算机有朝一日解决传统方法难以解决的问题的可能性。

已经提出了许多分子计算的模型，其中一些已经通过实验验证了可行性。

三~~变量~~可满足性问题是典型的NP完全问题。目前，NP完全问题已知最快的顺序求解算法需要指数时间。在Lipton证明了分子计算的并行性在求解这类问题具有独特优势之后，求解NP完全问题成为了测试DNA计算机性能的基准。Simth团队基于表面计算求解了四变量的可满足性问题（16种可能的真值赋值）。Yoshida和Suyamza还使用DNA程序执行广度优先搜索的策略解决了一个四变量实例。Sakamoto等使用DNA发夹结构解决了一个六变量问题（64种可能的真值赋值）。Landweber团队基于RNA解决了一个九变量可满足性问题的实例（512种可能的真值赋值）。本文求解了一个20个变量三可满足性问题的实例（1048576可能真值赋值）。

在本研究中，所采用的架构（模型，前面的architecture翻译为模型）与Roweis提出的粘贴模型有关。粘贴模型包含两个基本的计算操作：基于子序列的分离和粘贴操作。本研究只涉及分离操作，用聚丙烯酰胺凝胶填充玻璃模块，将寡核苷酸探针（单链DNA）固定在玻璃模块中实现分离。携带信息的DNA链通过电泳在模块中移动，与固定探针序列互补的DNA链则被捕获保留在模块中，不互补的DNA链则相对自由的通过模块。通过在高于探针-DNA信息链组成的双链体的解链温度下电泳，释放被捕获的链~~被释放~~。

使用电泳在凝胶填充的玻璃模块间传输DNA链，可以使计算机“干燥”并具有潜在的自动化能力。在分离过程中，共价键既不形成也不断裂，DNA链和玻璃模块可以多次用于计算。

布尔公式 本文计算的是20个变量24个子句3-CNF问题（如图1a）。为了使计算尽可能具有挑战性，问题被设计为只有唯一的解（如图1b）。仔细观察可以发现，这种设计赋予了一种迭代的语法结构。然而，这里的DNA计算没有使用这种结构，而是通过穷尽（exhaustively searched，摘要第二句翻译为详尽）了所有的220种可能的真值赋值，来寻找唯一满足条件的解。因此，可以合理的推断，任何20个变量24个子句3-CNF问题都可以容易地被解决。

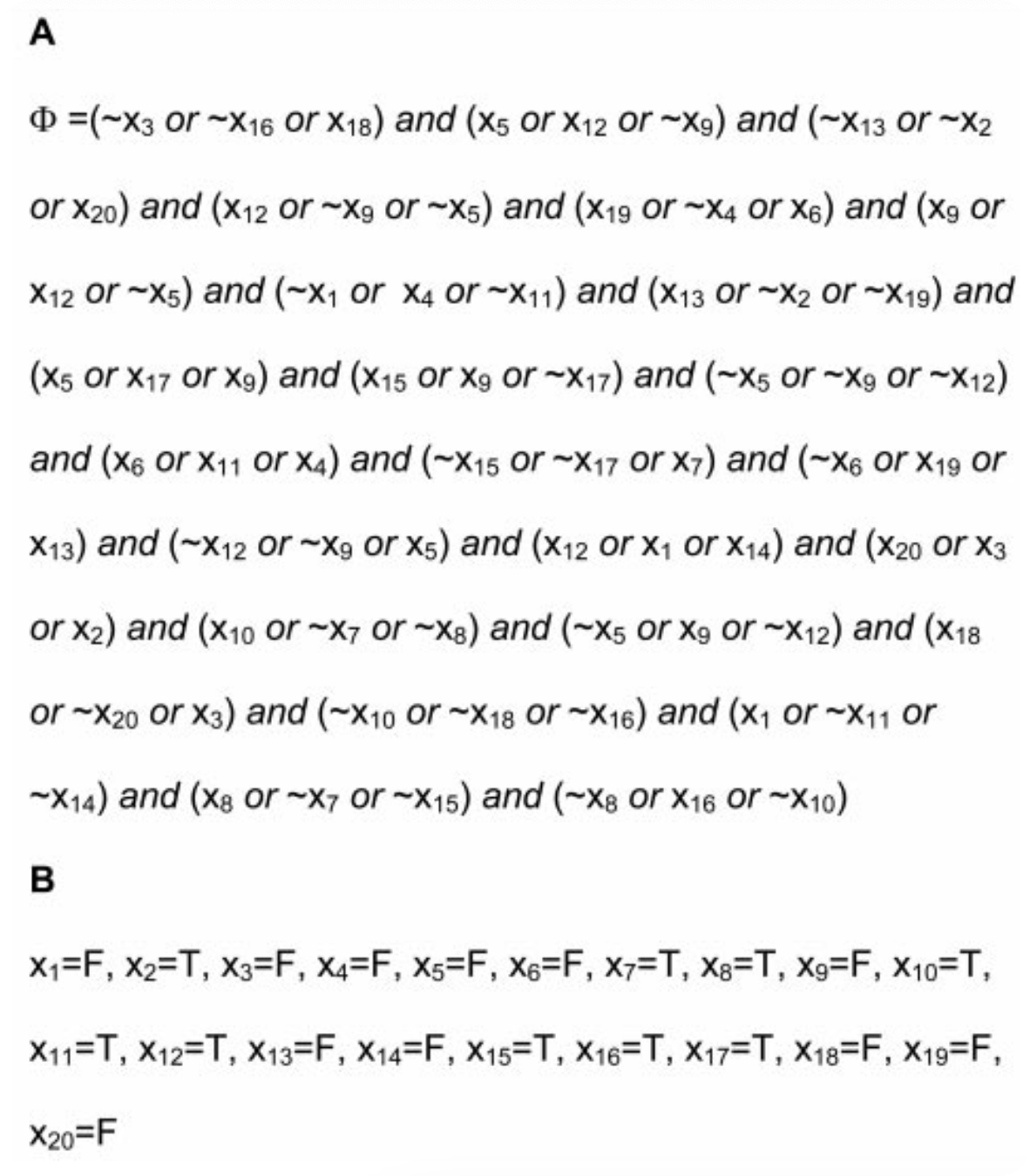


图1 计算问题 (a)20个变量的3可满足问题。“～”表示取反 (b) 唯一的成真赋值

库 使用Lipton编码表示了所有可能的真值赋值。对于20个变量中的每一个变量都用两个不同的15个碱基的“值序列”表示：一个表示“真”（T），，另一个表示“假”，。表示的Watson-Crick 互补序列。每一种真值赋值都由300个碱基的“库序列”表示。库序列由每一个变量的一个值序列的有序连接而成。具有库序列的单链DNA分子被称为“库链”。所有库链与互补链的集合被称为一个“完整库”。

对每一个值序列的互补的5‘端都用丙烯酸酯修饰用作分离操作的探针。

为了减少计算错误，需要设计序列（序列需要被设计为）阻止库内及库间的链杂交以及非预期的探针库链杂交。为了达到这个目标，计算机生成的序列需要满足之前报道的约束条件，特别是值序列中不能含“G”。

在自动化DNA合成仪上合成长分子可能是低效的。为了避免低效操作，先创建两个“半库”再组成完成库（full library，前文为完整库）。左半库用于到，右半库用于到。使用聚苯乙烯基固体载体在双柱ABI 392 DNA/RNA合成仪上以0.2 µmol的规模合成半文库。使用混合和分裂结合合成技术。简单的说，对于左半库，~~在~~分别在单独的柱中合成具有序列和的寡核苷酸。从合成仪中移除并打开两个柱子。将固体支撑珠（solid-support beads，前文的solid support为固体载体）混合在一起并分成两半，然后~~将~~分别放入不同的柱中。关闭柱子，然后分别按按照和序列重新开始合成。重复此过程，直到处理完所有的10个变量。右半库使用了类似的过程。

为了评估半库的简并度和用丙烯酸酯修饰的探针捕获半库的有效性，进行了凝胶捕获实验[web 图1]。对于40序列中的每一个序列，通过将相应的丙烯酸酯修饰探针添加到聚丙烯酰胺凝胶中，创建一个“捕获层”。通过电泳~~将~~，5‘端修饰磷酸基团的半库等分试样通过捕获层。与预期一致，对于每一个探针，半库的大约一半的链被捕获，而大约一半的链通过。这表明探针固定在了凝胶中并捕获了~~一半库链~~（half-library strands，半库链）。这也表明，半库链具有与探针完全互补的子序列，并且对于对于每一个变量，为“真”的半库链的数量与为“假”的半库链的数量大致相等。

为了进一步测试半库，在标准条件下，使用400 fmol的左半库作为模版，用引物组进行聚合酶链式反应（PCR）扩增35个循环。类似地，在标准条件下，使用400 fmol的右半库作为模版，用引物组进行聚合酶链式反应（PCR）扩增35个循环。凝胶电泳分析表明，在所有情况下都获得了预期长度的产物。证实了子序列，，，存在于左半库的预期位置，并且子序列，，，存在于右半库的预期位置。

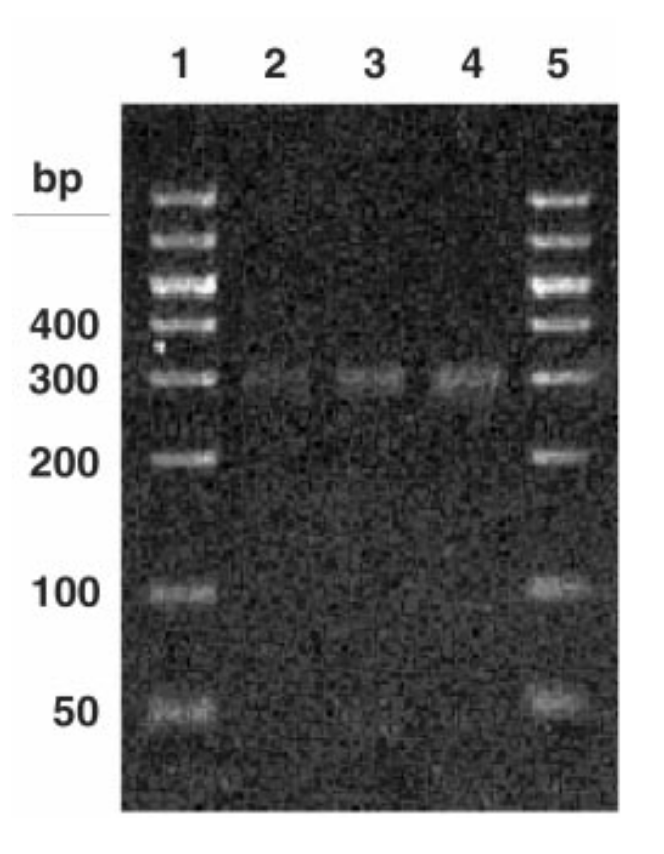


图2 全库（完整库）的构建。四个30 mer的寡核苷酸 “夹板 ”序列：、、、。每个半库10 pmol与4个夹板各2 pmol混合，加入终体积为25 ul的1XT4 DNA连接酶缓冲液中，室温孵育2 h (注:未加连接酶)。取1/2 ul 混合液，用引物组，在标准条件下进行PCR扩增35个循环。随后，从上述产物中取1 ul 的等分，使用上述引物组，在标准条件下进行35个循环的PCR扩增。将产物在1%琼脂糖凝胶上运行，使用Ultrafree-DA DNA提取试剂盒提取300bp的条带，然后将其混合以制备最终体积约 500ul 的原液。使用上述引物组合，使用2.5ul的原液取样作为模板，在标准条件下进行了另一轮PCR扩增，进行了15个循环。将得到的DNA进行了乙醇沉淀，并在75 ul 水中重新溶解。如图所示，完整库的取样在4%的琼脂糖凝胶上运行结果，1ul（泳道2）、2ul（泳道3）、3ul（泳道4）。泳道1和5为分子量标记。

使用类似于文献[26]中所述的聚合酶延伸方法，从两个半库创建完整库（300-mer）。随后，为了生成足够的计算所需的完整库，进行了两个阶段的PCR扩增。凝胶电泳结果显示仅有一条与最终产物对应的300bp条带（图2）。分光光度分析显示完整库为总共约750 pmol长300bp的DNA。值得注意的是，完整库由丙烯酸酯修饰的互补双链构成。~~以为~~（因为）所有库链都有相同的长度（300个碱基），所以他们在电泳过程中的运行速度相似。

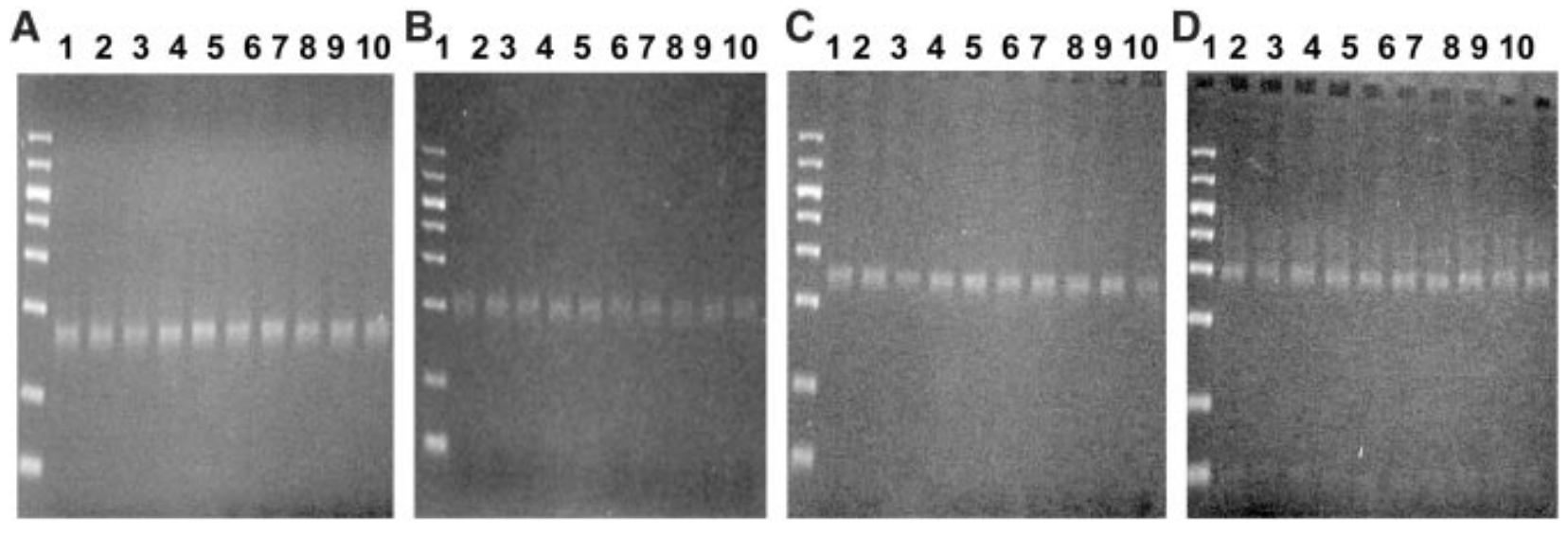


图3 全库（完整库）的分析。纯化后的完整库在标准条件下进行PCR扩增，15个循环。PCR产物在4%琼脂糖凝胶上进行分析。泳道1和2对应引物组，泳道3和4对应引物对，泳道5和6对应引物对，泳道7和8对应引物对，泳道9和10对应引物对，其中（A）k=11；（B）k=14；（C）k=17；（D）k=20。分子量标记位于每个凝胶的最左边泳道(如图2所示)

为了测试完成库（完整库），对任一k值，使用引物组 （）进行PCR扩增。所得产物的凝胶电泳分析显示预期长度条带（图~~三~~3）。

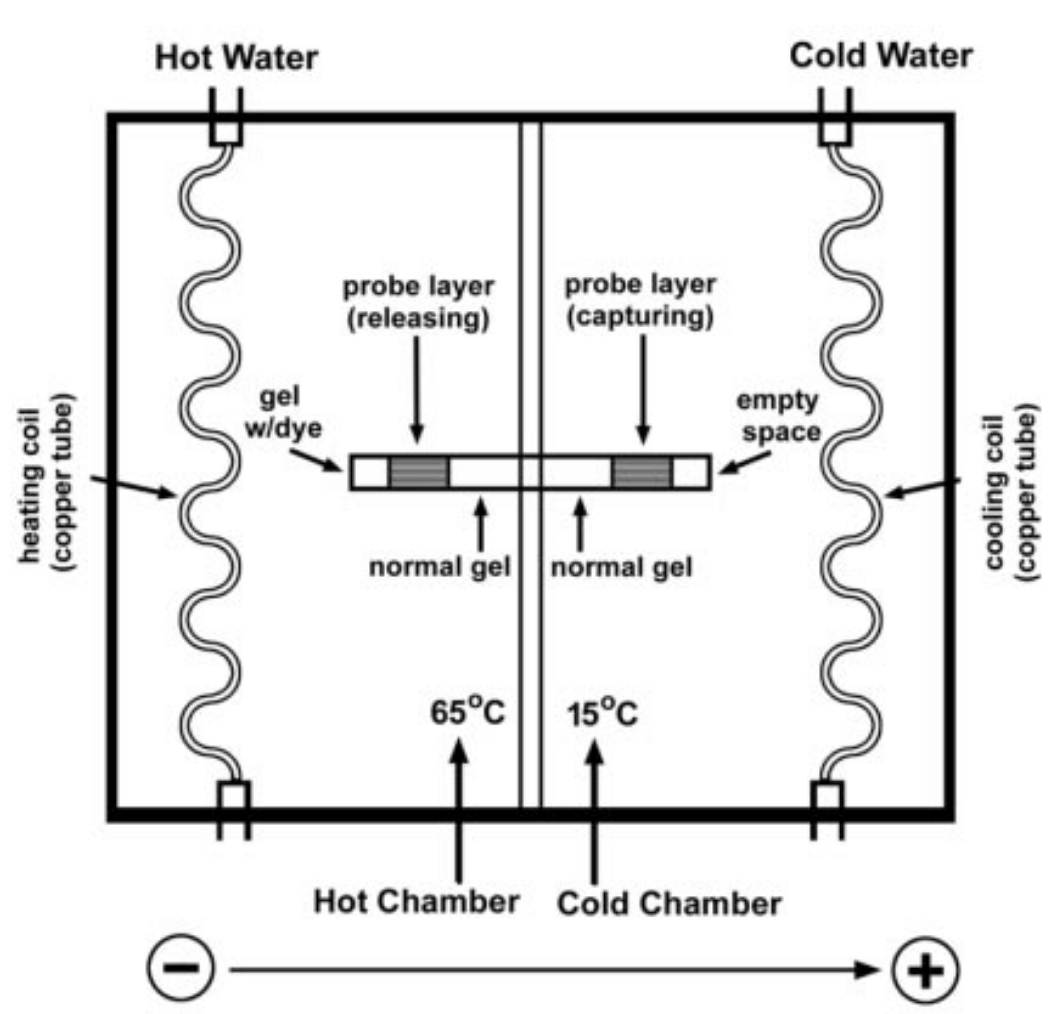


图4 计算机。电泳箱子由0.5 cm厚的有机玻璃制成，长30 cm，宽15 cm，高8 cm。箱体由有机玻璃隔板隔成等体积的热室和冷室。每个腔室通过塑料管连接到一个循环水浴中。循环水浴槽中的水通过铜管输送到腔室中，充当冷却/加热线圈。在每个腔室中插入了一根铂丝电极。对于𝜱 的24个子句中的每一个，准备了100 ul 的子句溶液，其中包含三种Acrydite修饰的探针，各15 uM，分别对应每个子句中的文字（如果子句中出现xk，则加入序列为的探针，如果子句中出现 ～xk，则加入序列为的探针）。例如，对于第一个子句（～x3 or ～x16 or x18），添加了序列为，和 的探针。对于每个子句溶液，在4.5 cm长的玻璃管中创建子句模块，外直径为0.5 cm，内直径为0.3 cm。在管底的3.2 cm处聚合了一层5%聚丙烯酰胺凝胶作为基层。在基层上方聚合了一层含有子句溶液的5 %聚丙烯酰胺凝胶作为探针层。如上所述制备了一个库模块，但使用了100 ul、500 pmol的完整库代替子句溶液。去往热室模块的末端用含有溴酚蓝和二甲苯蓝染料的5%聚丙烯酰胺凝胶进行堵塞。热室温度为65 °C，冷室温度为15 °C。电泳在12 V / cm下进行。染料在凝胶中的移动提供了一种监测过程和检测模块之间可能的泄漏的方法。当二甲苯蓝染料通过模块并进入冷缓冲溶液时，电泳停止，大约4小时。

计算机和计算方案 该计算机由一个带有热室和冷室的电泳箱、一个装有聚丙烯酰胺凝胶的玻璃“库模块”组成，其中含有共价结合的全库（完整库），对于的24个子句中的每一个，都有一个装有聚丙烯酰胺凝胶的玻璃“子句模块”，其中含有共价结合的探针，旨在仅捕获编码满足该子句的真值赋值的库链（图4）。计算方案如下：

第一步：将库模块插入电泳箱的热室，第一子句模块插入电泳箱的冷室。开始电泳。理论上，该步骤中，库模块中库链与丙烯酸酯修饰的探针链分开，并迁移到第一个子句模块中。编码满足第一子句真值赋值的库链在捕获层中被捕获，而编码不满足赋值的库链穿过捕获层并继续进入缓冲库。例如，具有序列、或的库链被保留在捕获层中，而具有序列、或穿过模块（捕获层）。

第二步：从电泳箱卸下两个模块。丢弃热室中的模块，清洗盒子并加入新的缓存液（buffer，前文为缓冲）。将冷室中的模块插入热室中，下一个子句模块~~中~~插入冷室中。开始电泳。理论上，在步骤二中，热室中子句模块中库链与丙酸乙酯修饰的探针链解链（melt off，前文为分开），并迁移到冷室中的下一个子句模块中。编码满足这一子句真值赋值的库链在捕获层中被捕获，而编码不满足赋值的库链穿过捕获层并继续进入缓冲库。

第三步：对剩余22个子句依次重复步骤二。理论上，步骤三结束时，最终的子句模块（第24个子句模块）将包含所有24个子句模块中捕获的那些库链，并因此满足每一个子句的真值赋值和式的真值赋值。

第四步：从最后一个子句模块中提取解链，进行PCR扩增，并“读出”解。

确定解 从最后一个子句模块中提取凝胶，并在65°C的1mL水中浸泡过夜，析出库链。将库链冻干，在200 µL水中复溶，脱盐，并在45 µL水中回收，形成解预备集。

为了求解变量和变量的真值，用引物组分别扩增1 µL解预备集的10倍、20倍、30倍、40倍、50倍、60倍和100倍稀释液。10倍、20倍、30倍、40倍、50倍稀释的PCR产物的凝胶分析显示，除了引物外其他引物没有条带。这些引物集（primer sets，前文为引物组）仅有一个对应300bp的条带，因此，变量的解为*F*，变量的解为*F*。60倍和100倍稀释液的PCR产物分析显示没有任何条带。

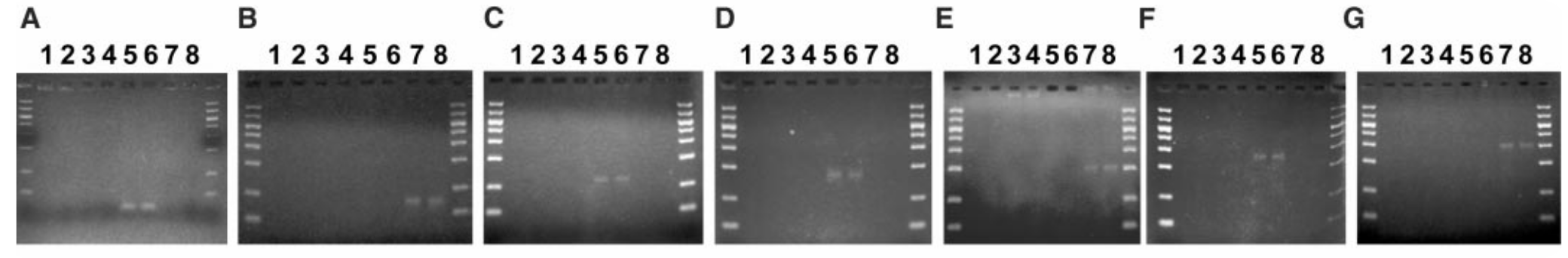


图5 解的读出。取1 ul的解预备集的50倍稀释液，在标准条件下进行25个循环的PCR扩增。PCR产物在4 %琼脂糖凝胶上分析。泳道1和2对应于引物对，泳道3和4对应于引物对，泳道5和6对应于引物，泳道7和8对应于引物对，其中（A）k=2，（B）k=5，（C）k=8，（D）k=11，（E）k=14，（F）k=17，（G）k=20。分子量标记位于每个凝胶的最左边和最右边泳道(如图2所示)

为了求变量的真值，并作为变量真值的冗余测试，使用引物集（引物组）（）分别扩增1 µL解预备集的50倍稀释液。每一个*k*值对应的凝胶电泳显示~~有切~~（有且）仅有一种引物组合存在~~了~~预期长度的条带（图5）。根据条带信息，得到每个变量的真值。实验得到的每个变量的真值即为式唯一的解。

捕获-释放效率 对正确链，即序列满足式唯一真赋值的链，进行了以下分析。因为是取45µL 解预备集的50倍稀释液中的1 µL进行PCR得到了正确的解。因此，解预备集中（（it is probable that）可能）至少包含了条正确链。由于计算开始使用的是大约500 pmol（个分子）的完整库，加之大约每220个分子中有一条正确链，因此整个计算中正确的解存在的概率至少为。因为计算包括了24个捕获-释放步骤，（（it follows that on average,）因此平均而言，）正确链在单个捕获-释放阶段存活（surviving，前文为存在）的概率至少为。上述分析建立在PCR能够检测到单个模版链之上，一个更实际的假设可能是，仅在25个周期后获得阳性信号至少需要5000条模版链（at least 5000 templates were required to produce a positive signal after just 25 cycles，在25个循环后，至少需要5000条模板链才能获得阳性信号。）。基于这个假设，正确链在单个捕获-释放阶段中存活的平均概率至少为。

对不正确链，即不能使得式为真的链，分析如下：假设至少需要5000个模版链才能在25个PCR循环后产生阳性信号。取1 µL解预备集（共45 µL）为底物进行PCR显示了错误真值赋值的条带。然而，取1 µL解预备集的10倍稀释液为底物进行PCR，未发现~~于~~错误的真值赋值的条带。因此，解预备集中最多可能有条错误链。可以合理地解释（assume，假设），所有不正确链都是同一种不正确的真值赋值方案。这种链称为“1-错误链”。对每一条“1-错误链”，通常都有一个不应退火的捕获层。在开始计算时，大约有条“1-错误链”。因此，在关键捕获-释放步骤意外保留“1-错误链”的概率最多为。

该分析表明，基于上述方法使用500 pmol的全库（完整库）可以解决大约30个变量的3可满足问题。计算过程中周期性的PCR扩增可能会扩大这种方法求解~~2~~3可满足性问题的规模。

结论 数千年来，人类一直试图制造设备来增强固有的计算能力。诸如算盘、加法器和制表机等机械设备都是重要的进步。然而，直到大约60年前电子设备，特别是电子计算机（的出现），计算能力得到了质的飞跃，相对困难的问题才得以解决。现在看来，分子装置有望取得进一步的突破。

在我们的研究中，使用了一种简单的方法。以DNA沃森-克里克碱基互补配对和解链作为“操作”解决了20个变量3可满足问题的一个实例。尽管计算理论可以预测，但这种基本的分子相互作用能够维持如此复杂的计算，这仍然是非同寻常的。我们的实现是粘贴模型的简化版本。~~我们没有使用贴纸，因此，我们的库链表现得有固定的记忆。~~

我们没有实现贴纸，因此我们的库链表现得像固定的存储器。如果按照最初设想，库链将充当更强大的写一次（单次写入，write-once）存储器。最近的研究表明，DNA“链入侵”可能为从库链中特异性去除粘附物（stickers，前文为贴纸）提供了一种方法。这可能会使库链成为非常强大的读写存储器。进一步调查这种可能性似乎是值得的。

尽管我们和其他研究人员一样取得了成功，但是在缺乏技术突破的情况下，我们不能盲目相信可以创造一台能够在经典计算问题上与电子计算机抗衡的分子计算机。然而，分子计算机可以在更广泛的背景下考虑。它们可能在特殊环境中有用，例如，需要极高的能量效率或非凡的信息密度。它们可能为控制化学/生物系统提供了一种急需的手段，就像电子计算机为控制电气/机械系统提供了一种手段一样。它们确实为生物和计算思想的整合提供了一个焦点，这可能会带来更广泛的应用，例如在DNA自组装方面有前景的工作。它们为我们提供了电子计算机的替代品，研究它们最终可能会引领我们走向真正的“未来计算机”。

最重要的是，DNA计算机，如这里展示的，表明生物分子可用于非生物目的。为此，这些分子代表了30亿年进化中未开发的遗产，并且在其进一步探索中具有巨大的潜力。