

DNA 与分子计算机

DNA and Molecular Computer

黄 威

(清华大学生物科学与技术系, 研究生 北京 100084)

刘进元

(清华大学生物科学与技术系, 教授 北京 100084)

自从 1946 年人们研制出世界上第一台电子管计算机以来, 计算机技术得到迅猛发展。计算机选用的部件从电子管到晶体管, 到集成电路、超大规模集成电路; 运算速度从 5 000 次/秒提高到上亿次/秒。计算机的“心脏”是中央处理器—CPU 芯片, CPU 芯片集成的晶体管数量越多, 计算机的功能越强, 同时对硅芯片加工的工艺要求越高。当今, CPU 芯片集成的晶体管数量已达数百万, 芯片加工工艺的线宽已达 $0.35\mu\text{m}$ 。由于半导体工艺加工线宽的极限是 $0.07\mu\text{m}$ 和晶体管相互的散热干扰, 导致芯片集成的晶体管数量不能无限增加, 因而预计在下个世纪, 传统电子计算机的发展将达到极限, 人们不得不考虑下一代计算机的发展。

DNA 是一种由很多单体组成的线性大分子, 它的基本结构是被称为核苷酸的重复单元, 其排列顺序蕴藏着生物体的全部遗传信息。特定的 DNA 分子所特有的核苷酸排列方式决定了它所携带的遗传信息及行使的生物功能。具体地说, 遗传的本质是 DNA 分子的基本结构所携带的遗传信息经转录传递给 mRNA 分子, 再以此信息为指令合成特异的蛋白质而行使生物功能。可以认为, 生物遗传和生命活动是由分子机器来实现的。

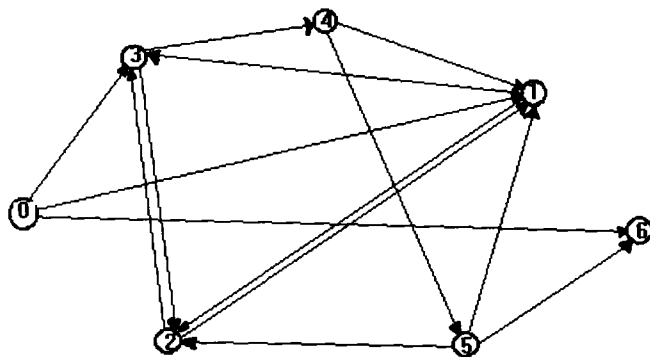
根据 DNA 的内在特性, 科学家提出了分子计算机的设想: DNA 分子的碱基序列相当于存储的数据, DNA 分子之间可以在某种酶的作用下发生特定的生化反应, 从一种序列变为另一种序列, 反应前的 DNA 序列可以作为输入的数据, 反应后的产物序列可以作为运算结果。相对于传统电子计算机, 分子计算机具有巨大的潜在优势, 是下一代计算机的发展方向之一。

一、分子计算机的产生

1994 年, 美国南加州大学教授 Leonard M. Adleman 在《科学》杂志上发表论文, 首次提出分子

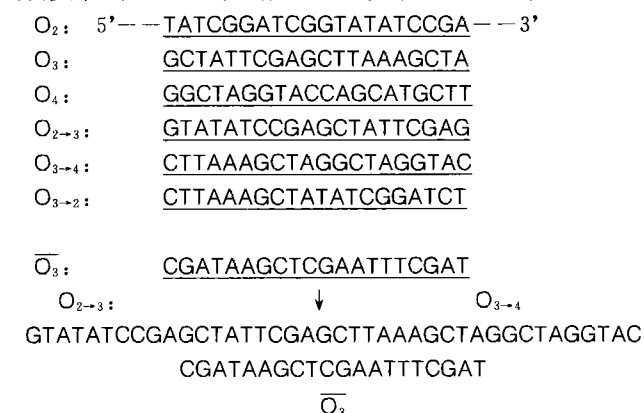
计算机的设想, 而且用构建的分子计算机解决了著名的汉密尔顿(Hamiltonian)路径问题。

汉密尔顿路径是一个经典的计算理论上的 NP 复杂性问题, 许多实际的图论计算问题都可以归结为汉密尔顿路径问题。根据算法解决问题所用时间与问题维数的关系, 计算理论将问题复杂性分为 P 和 NP。P 问题是指运算时间由一个限定次数的多项式决定。NP 问题是指运算时间由一个不定次数的多项式决定, 随着问题维数的增加, 多项式的次数随之增大, 算法的计算量也大大增加以至于超过常规电子计算机的计算能力。因此, NP 问题被认为是难以解决的。具体来说, 汉密尔顿路径问题是指在一个有向图中, 对于两个顶点 V_{in} 和 V_{out} , 如果存在一条开始于 V_{in} , 结束于 V_{out} 的单向路径, 且此路径通过其它所有的顶点有且仅有一次, 则称存在这条汉密尔顿路径。例如下图是一个由 7 个顶点和 13 条边组成的汉密尔顿图, 当 $V_{in}=0$ 且 $V_{out}=6$ 时, 它存在唯一的汉密尔顿路径, 即 $0 \rightarrow 1 \rightarrow 2 \rightarrow 3 \rightarrow 4 \rightarrow 5 \rightarrow 6$ 。而解决具有 n 个顶点的汉密尔顿路径问题的经典算法为: 先根据汉密尔顿图, 构造出所有的任意路径; 然后初次筛选, 保留开始于 V_{in} , 结束于 V_{out} 的路径;



再次筛选,保留只经过 n 个顶点的路径;再次筛选,保留通过所有顶点至少一次的路径;最后如果存在此路径,则结论为肯定,反之为否定。

那么如何利用分子计算机原理来解决汉密尔顿路径问题呢?以上图为例,Adleman 教授是这样来构建分子计算机而实现此算法的。对上图中的任意一个顶点 i ($i=0,1,2,3,4,5,6$),用 20 个任意核苷酸组成的 DNA 片段 O_i 来表示,对起始于顶点 i 终止于顶点 j 的边 $i \rightarrow j$ 用 O_i 的 3' 端的 10 个核苷酸序列($i=0$ 时,用整段 O_0)连上 O_j 的 5' 端的 10 个核苷酸序列($i=6$ 时,用整段 O_6)来表示。例如:



$\overline{O_i}$ 是 O_i 的互补 DNA 序列,在连接反应中,作为接头连接两段 DNA;并在往返中确定连接的方向性,如 $O_{2 \rightarrow 3}$ 与 $O_{3 \rightarrow 2}$ 的核苷酸序列就不相同。具体步骤:

(1)输入:对每一个顶点 i ($i=0$ 和 $i=6$ 除外)和每一条边 $i \rightarrow j$,加入 50pmol 的 $\overline{O_i}$ 和 50pmol 的 $O_{i \rightarrow j}$,混合后发生连接酶作用下的连接反应。 $\overline{O_i}$ 帮助相关的连接。最终形成的 DNA 分子就编码了所有的任意路径。

(2)筛选:选用 O_0 和 $\overline{O_6}$ 作为引物,对产物进行 PCR 扩增,筛选出编码 $V_{in}=0$ 和 $V_{out}=6$ 的 DNA 分子;然后进行琼脂糖凝胶电泳,筛选出 140bp 长度的 DNA 片段(即编码通过 7 个顶点的路径);再用生物素-抗生物素磁珠系统(biotinavidin magnetic beads system)进行亲和纯化,将 140bp 的 DNA 片段变性得到单链 DNA,单链 DNA 与 $\overline{O_1}$ (结合在磁珠上)混合培养。只有含有 $\overline{O_1}$ (即编码路径通过顶点 1 至少一次)的 DNA 片段才能经退火连接 O_1 而被保留。此后重复以上步骤,使用 $\overline{O_2}, \overline{O_3}, \overline{O_4}, \overline{O_5}$, 筛选出通过所有顶点的 DNA 分子。

(3)输出:采用分级 PCR(graduated PCR)扩增方法输出结果,即选用 O_0 作为 avidin magnetic beads system)进行亲和纯化,将 140bp 的 DNA 片段变性得到右端引物, $\overline{O_i}$ ($i=1,2,3,4,5,6$) 作为左端引物,对最终产物分别进行 6 次 PCR 扩增反应,扩增产物再经凝胶电泳鉴定。如存在此片段,则可得所求路径;反之则说明不存在此汉密尔顿路径。

结果本例中经分级 PCR 扩增得到 6 条分别为 40bp, 60bp, 80bp, 100bp, 120bp, 140bp 的电泳带。第一次 PCR 扩增产物为 40bp:代表顶点 0 到顶点 1 的路径长度为 40bp,说明顶点 0 到顶点 1 是直接连接,即编码路径 $0 \rightarrow 1$;第二次 PCR 扩增产物为 60bp:代表顶点 0 到顶点 2 的路径长度为 40bp,说明顶点 0 到顶点 2 之间存在一个其它顶点,即编码路径 $0 \rightarrow 1 \rightarrow 2$ 。以此类推最终可以得出汉密尔顿路径为: $0 \rightarrow 1 \rightarrow 2 \rightarrow 3 \rightarrow 4 \rightarrow 5 \rightarrow 6$ 。

Adleman 就这样构建出世界上第一台分子计算机,简单巧妙地解决了汉密尔顿路径问题。用分子计算机来解决类似汉密尔顿问题主要有 3 点优势。(1)本例中连接反应的规模远超过所需,对每一条边而言,大约有 3×10^{13} 个拷贝的 DNA 片段加入到反应中,这就保证生成足够的汉密尔顿路径。事实上,对本例而言 attomole 量级的 DNA 已经足够反应用量。当问题规模扩大,顶点增多,则反应物的量相应增加。(2)选用 20 个核苷酸表示顶点的原因是 20 个核苷酸有 4^{20} 种排列可能,避免不同顶点具有相识的 DNA 片段,且 DNA 片段在室温时较稳定。(3)对于汉密尔顿图,顶点数目增加时,可能的汉密尔顿路径也大大增加,问题的计算量随之增加,当顶点数目超过一定限度,即使最快的电子计算机也无能为力;而 DNA 溶液具有巨大的并行计算能力,能够探索所有可能的汉密尔顿路径,有可能解决顶点数目更多的汉密尔顿路径问题。

二、分子计算机的意义及其特点

在理论上,分子计算机的产生,不但打破了传统意义上的计算机概念,提出了计算机发展的新思路,而且有助于揭示生命的本质。分子计算实际上反映出简单的生物系统具有巨大的计算能力。Adleman 实验采用的仅仅是连接反应的机制,生物体内还存在其它多种多样的酶反应过程,如转录翻译、基因调控等。任何一个生物学功能都是通过一组生化酶反应来实现的,都是在特定的机制作用下的物理和化学过程,本质上都体现了一种严格的计算和算法特征;那么一维的 DNA 序列是如何实现和控制复杂的生命呢?分子计算的实例暗示出, DNA 作为遗传信息的载体,它既是编码数据的数据库,又是编码算法的程序库。程序和数据具有相同的格式。生命就是在 DNA 控制下的程序运行过程。

在实际应用上,分子计算机也具有自身的潜在优势。传统电子计算机用磁介质存储信息,单位时间内运算速度很快,但参与并行运算的处理器数目很少。与其相比,分子计算机具有 4 个特点:

(1)运算速度极快 普通的 PC 计算机运算速度为 10^6 次/秒,目前最快的超级电子计算机速度为

10^{12} 次/秒。而对于分子计算机,如果把两个DNA分子连接视为一次操作,又假定 4×10^{14} 个边的DNA片段有一半发生了连接反应,则分子计算机的运算速度为 10^{14} 次/秒。事实上,如果增加参与连接反应的DNA量时,甚至可以达到 10^{20} 次/秒或更高。这种速度远远高于现有的超级计算机。

(2)消耗的能量极少 在一次连接操作中 $\text{ATP} \rightarrow \text{AMP} + \text{焦磷酸}$ 的 $\Delta G = -8\text{Kcal/mol}$,则1焦耳的能量可以完成 2×10^{19} 次运算;而对于超级电子计算机,1焦耳的能量至多能完成 10^9 次运算。相比之下,即使考虑到核苷酸片段合成和PCR扩增的能量消耗,分子计算机也具有巨大的节能优势。

(3)能提供巨大的信息存储容量 使用DNA分子存储信息密度可以达到1比特/ 10^{-9}nm^3 ,而常规媒体只能达到1比特/ 10^{12}nm^3 。1立方米的DNA溶液可存储1万亿亿位的数据,超过目前世界上所有计算机的存储容量。

(4)可以真正实现并行工作 传统电子计算机主要是串行工作,而分子计算机可视为多CPU的并行工作,可以实现现有计算机无法真正实现的模糊推理和神经网络运算功能,是智能计算机的突破口之一。对于分子计算机,一个DNA分子相当于一个CPU,在1molDNA溶液中就含有 10^{23} 个分子,则可以实现 10^{23} 量级的并行计算。

三、分子计算机的发展

今天的电子计算机经历了半个世纪的发展,已经走向成熟;而分子计算机仅有两年的发展时间,正处于探索时期。它能否应用于现实的计算问题还有待研究,但至少在传统计算机难以解决的一些特

殊问题,如NP问题上,分子计算机显示出巨大的潜在优势。当然,在它的发展道路上还面临许多挑战:在生物学和化学方面科学家需要更好地理解和掌握细胞和分子反应机制,以期实现分子计算机的算法;在计算机科学和数学方面科学家需要找到合适的课题和有效算法,方可运用分子计算机;在物理学和工程学方面科学家需要发现更好的结果输出方法,才能构建更大规模和更可靠的分子计算机。

目前,分子计算机已经成为现代科学的热点问题之一,许多科学家都投身于此领域,不断取得惊人的成果。在Adleman用分子计算机解决了汉密尔顿路径问题后;Lipton又将此算法推广用于解决任何NP问题;而Rooß和Wagner则证明对于任何复杂性 $P^{\text{NP}} = \Delta_2^P$ 的问题都能用此算法解决。分子计算机正处于积极探索时期,亟需大力推动。

综上所述,分子计算机将扩展计算机发展的思路,加深对计算机和生命本质的理解。计算机本质上只是遵从物理定律的物理装置。未来的计算机或许将不再是电路板,而只是装在试管中的DNA。

参考文献

- [1]Leonard M. Adleman, Science Vol. 266, P1021~1024
- [2]Michal Linial, Science Vol. 268, P481~484
- [3]R. Lipton, Science Vol. 268, P542
- [4]David K. Gifford, Science Vol. 266, P993
- [5]S. Cheng, S. Y. Cheng, Nature 369, 684(1994)
- [6]邓少平, 科学(中文版), 1996. 10, P6~8
- [7]Diana Rooß and Klaus W. Wagner, Information and Computation, Vol. 131, P95~109
- [8]Ostrovsky—B Annual review of Biophysics & Biomolecular structure 1995, Vol 24, P239~267

(责任编辑 王宏章)

科技动态

美国研制用液氮推动的汽车

据英国《新科学家》1997年8月报道,美国西雅图华盛顿大学的荣誉教授阿贝·赫茨伯格领导的一个科研小组改装了一辆老式邮政车。它的发动机像老式蒸汽机一样工作,但其中的蒸汽不是用水,而是用液氮。传统的燃煤蒸汽机用锅炉将水烧成高压蒸汽,用来推动活塞式发动机发电。而赫茨伯格的发动机用的是液氮,它在 -196°C 就会沸腾,因此不需要外部的燃料源。液氮只要经过一个空气热变换器就能被气化,气化的液氮产生足够的压力就能推动活塞式发动机。

排气管排出的是纯净的低温氮气,不会产生任何污染,而且有助于消除温室效应。赫茨伯格说,生产液氮当然也要消耗能源,但生产成本较低。汽车每公里使用液氮的成本可能小于电动汽车的成本。液氮的另一个优点是,在冷却时,空气中的二氧化碳在氮变成液体之前早就冷凝成固体了,因此可以在生产液氮时从空

气中分离出这种温室气体,为防止全球变暖作出贡献。

赫茨伯格的实验性液氮发动机汽车,用180升液氮可以行驶约160公里。如果将液氮箱的容积增加到360升,估计可行驶400公里,和电动汽车行驶的距离相当。赫茨伯格在1997年8月上旬在圣迭戈举行的汽车工程学会讨论会上宣布了他们的研究成果。但有人怀疑这种汽车是否实用,因为液氮箱占用的体积太大了。比如,美国普林斯顿大学能源和环境研究中心的罗伯特·威廉斯说,如果一辆汽车的液氮箱就占了360升的空间,那这辆汽车大部分就是发动机和液氮箱的位置了,而且还可能出现液氮引起的冻伤,形成新的公害。

尽管有不同的意见,但赫茨伯格的设想仍不失为一种新颖的构想。

(刘光曙)