

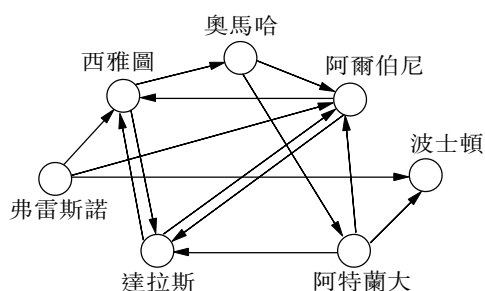
DNA計算與生物數學

• 孟大志 陸薇

二十一世紀是生物學的世紀，以「人類基因組計劃」拉開序幕。這項計劃將在2003年前完成人類脫氧核糖核酸 (DNA) 全部序列的精確圖，從此生物科學將進入破解這部天書的「後基因組時代」。破解這部天書，就是要讀懂DNA序列上怎樣編碼產生生命的全部資訊，而從數學角度說，就是理解生命是如何通過DNA序列的資訊而被「計算」出來的。從這個角度來看，DNA計算是非常值得關注的新學科。DNA計算誕生於1994年11月，作為新學科，它的意義在於DNA可以作為計算的介質而用於解決數學問題，這是真正意義上的生物數學——生物是計算工具，而人類的數學問題是計算的目的。這完全不同於以往科學文獻中那種以數學為工具、以生物學問題為目的的「生物數學」或「數學生物學」概念。DNA計算之所以受到科學界關注，是因為它預示着嶄新的經濟可能：創造本質上並行的DNA電腦，其運算速度、能耗效率及存貯密度都具有極為可觀的優越性。

一 新的計算概念：阿德勒曼的構想與實驗

1994年11月，阿德勒曼 (Leonard M. Adleman) 在《科學》(Science) 雜誌上發表論文^①，第一次提出人類用生物材料作計算物質以解決數學問題的思路，即所謂「哈密爾頓路徑問題」(Hamiltonian Path Problem, 簡稱HPP)。其要義是：在由有限個頂點與連接頂點之間、一個有方向的路徑所組成的網路圖中，是否存在一條可以走遍全部頂點、且每點只經過一次的路徑？郵遞員的最佳投遞路徑，或者飛機航線的確定等都可以歸結為哈密爾頓路徑問題。



以飛機航線的選擇可以形象地說明哈密爾頓路徑問題。圖中每一個頂點代表一個機場，箭頭表示飛機從一個機場到另一個機場的方向。要求飛機從起點(設定弗雷斯諾機場)到終點(設定波士頓機場)，必須經過圖中的每一個機場一次而且僅僅一次，確定符合這個要求的航線，用圖論的語言來說，就是確定「哈密爾頓路徑」。如果機場(頂點)的數量很少，可用窮盡搜索所有可能航線的方法求解。但頂點的數量如果增多，需要搜索的路徑會以指數級別增加，構成一個數學難題。

(資料來源：<http://www.ams.org/new-in-math/cover/dna-abc2.html>)

阿德勒曼實驗是為了解決哈密爾頓路徑問題，這是一個數學難題，屬於NP完全問題 (NP-Complete Problem)。所謂NP問題是指在非多項式計算時間內能計算完成的問題，也稱複雜性問題。而NP完全問題是這類問題中具有典型意義的一種。在理論上，NP完全問題可以用「窮盡式搜索」(exhaustive search) 來解決，而且已經開發了各種求解的演算法。但對一般的有向圖 (oriented graph) 來說，窮盡式搜索具有指數級的複雜性，在圖靈機計算系統中不可能找到一個有效的演算法 (即運行多項式時間的演算法)。阿德勒曼在實驗中，用DNA串的大規模並行克服了搜索的非確定性困難，從而以DNA計算進行窮盡式搜索，解決了哈密爾頓路徑問題②。

阿德勒曼的實驗基於如下簡單事實：1、生物體異常複雜的結構是對由DNA序列表示的初始資訊執行一系列簡單操作的結果；2、可計算函數可以通過一系列基本的簡單函數的計算而得到。洞察兩者之間的相似性，正是DNA計算誕生的基礎。

DNA計算是通過對DNA分子的操作來實現的。所謂「操作」，一方面是來自DNA在「遺傳」過程中存在的「自然」操作，它是DNA自身進行計算的工具；而另一方面也來自「遺傳工程」，因為操作DNA分子的能力是遺傳工程的關鍵。自70年代以來，遺傳工程的發展使人類已經積累了豐富的處理 (操作) DNA分子的方法，這些方法加入到DNA自身使用的「自然操作」集合中，形成了在DNA計算中操作DNA的基本工具庫。目前，對DNA分子的操作，既有物理的，也有化學的。物理操作實質上是調控生化反應的外部條件，例如溫度、酸鹼度等等。此外是來自各種生化試驗手段，尤其是通過各種酶的操作，是DNA計算的主要工具。通過物理和生化條件控制，可以實現對DNA的分離、結合、延長、縮短、剪切、長度測量和序列解讀等等「操作」③。

阿德勒曼不僅洞察到生物過程和數學計算過程之間的高度相似性，而且發現了將生化技術用於對DNA的操作去類比數學計算的可能：用DNA表示數學問題的初始資訊，酶用於類比簡單的計算，從而展現了以生物工具求解數學問題的可能。這項工作也是第一次用DNA計算解決一個數學難題。

二 DNA計算的完備性、通用性與自裝配模型

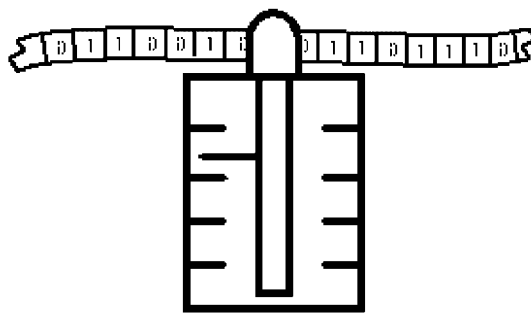
阿德勒曼的開創性工作還啟發人們思考以下兩個基本問題：DNA計算可以解決哪些數學問題？在理論上是否存在可編程的DNA電腦④？

實際上，自1994年以後，用DNA計算可以解決的數學問題的清單迅速增長。例如利普頓 (Richard J. Lipton) 描述了解決另一個NP完全問題：可滿足性問題⑤。此外符號決定性問題、道路染色問題、矩陣乘法、加法以及超標量電腦代數問題等等。特別是文獻給出一個「分子計算程式」⑥，它可以破解美國政府的資料加密標準 (DES)。在這種發展形勢下，人們自然要問：DNA計算可以解決任何問題嗎？這就是DNA計算的完備性問題。同時，從製造電腦的角度，自然會

提出：至少從理論上是否存在通用的DNA計算系統，是否能夠實現可編程的DNA計算？這就是DNA計算的通用性問題。從DNA計算誕生的第一年起，人們就關注這些基本問題。

與電子電腦不同，DNA計算程式是以含初始資訊的DNA鏈的試管作為輸入，輸出是另一些含結果資訊的DNA鏈的試管，計算過程是由含DNA鏈的試管的序列組成。這種計算操作導致了DNA計算獨特的編程。完備性和通用性問題的討論也是基於這種計算模式。由於電子電腦的堅實理論基礎，多數研究者都將DNA計算模型與現代通用計算模型——圖靈機進行比較。

由於丘奇—圖靈 (Church-Turing) 論斷被廣泛接受，人們確信任何實際的電腦的能力都不會超過圖靈機，因此將DNA計算與圖靈機作比較，意義就更加明確。通過專門處理符號系統的形式語言，人們已經證明：1、DNA計算模型可以計算所有圖靈機計算的函數，也就是說，DNA計算能夠完成目前電腦可以完成的所有工作，這就是其完備性證明^⑦。值得注意的是，相反的結論未必成立，也就是說DNA的計算能力並不一定受到丘奇—圖靈論斷的限制，因為關於它的能力的討論並不一定要在圖靈機的理论框架中討論；2、從計算理論上證明了可編程的DNA電腦是存



英國邏輯學家圖靈 (Alan Turing) 1936年提出了著名的「圖靈機模型」，包括三個核心部分：一條帶子、一個讀寫頭、一個控制裝置。帶子分成許多小格，每小格儲存一位數碼，以每次一格左右移動，讀寫頭受制於控制裝置，可以讀小格內的數碼符號，或者在帶子上寫符號。圖靈機可以把程式和資料都以數碼的形式存儲在帶子上，即「存儲程式」型，程式能把高階語言程式譯成機器語言程式，這就是「通用圖靈機」原理。圖靈機模型並不是製造具體機器的方案圖紙，而是提出了理想電腦原理的邏輯構造，後來成為通用數位電腦的數學模型。

(資料來源：<http://www.qubit.org/intros/compSteane/qcintro.html>)

在的。這一結論首先是用剪接系統為例證明的^⑧。事實上，各種生物操作的組合可以建立各種DNA計算模型，讀者可看比較詳細的專著^⑨。

但是與圖靈機的比較並不是必須的，一些作者認為DNA計算具有自己獨特的電腦制，尤其是它的極度並行的計算特性，因此不應該把DNA的計算模型納入傳統的電子電腦的框架中，而應該理解新的計算定義^⑩。然而即使在圖靈機的框架中，DNA計算模型的計算能力對比現代電子電腦，無論在速度、能源利用率及存貯資訊的經濟性方面均有十分誘人的潛力。例如在阿德勒曼模型中，每秒可執行 1.2×10^{18} 次操作，比當前最快的超級電腦要高出120萬倍；現有的超級電腦利用每焦耳能量可執行 10^9 次操作，而DNA電腦將可執行 2×10^{19} 次，即DNA電腦的能源效率要高出 10^{10} 倍；DNA分子中資訊存貯密度可達1bit/立方納米，是現有存貯介質的存貯密度的 10^{12} 倍，等等。

DNA計算的特殊性還在於其自裝配模型。自然界存在着許多自裝配過程，如原子反應生成分子、分子反應生成超分子或晶體等的過程都是自裝配。自裝

配是指相對簡單的分子在一定條件下，自動組裝生成較複雜分子的過程。實際上，自裝配就是自然界的電腦制。阿德勒曼利用這種電腦制解決了數學中的具體問題。DNA的計算模型是DNA分子的自裝配模型，它是指：在一定的溫度、濃度、酸鹼度以及特定的酶作用的反應條件下，一些帶有輸入資訊的DNA分子根據Watson-Crick互補配對原則，自組織生成新的、帶有輸出資訊的DNA分子的過程。如果用外部操作來有目的的控制反應條件，將分子的集合與操作組合相結合，就可以對應不同的計算模型，例如剪接系統、插入—刪除系統、黏接系統、試管模型、記憶體模型、PAM模型等等^①。就可以解決實際問題，實現DNA分子計算。

三 DNA電腦的可行性與技術困難

自1994年阿德勒曼的開創性文章發表以來，DNA計算研究的主要目標顯然就是為了製造真正的DNA電腦。儘管甚麼是DNA電腦這一概念還沒有統一，甚至甚麼是DNA計算都沒有明確定義，但是作為計算裝置的研究，既然對於某些可實現的計算模型與圖靈機比較有了相對的完備性與通用性，人們自然地以現代電子電腦的運作模式來構想DNA電腦可能採取的模式。在這些模式下，從當前的理論與實驗到實用的DNA電腦還有許多未解決的問題，可以歸納為以下三個方面：

- 1、生化技術對DNA計算操作的有效性和可靠性的影響。
- 2、怎樣解決大規模問題的誤差和DNA資源消耗？
- 3、甚麼樣的問題至少在理論上更適合於DNA計算？

在目前計算模型的研究中，多數通過生化實驗的實現，經過對實驗的分析，幾乎每項實驗操作都存在潛在的困難有待實驗科學家去改進。例如，生化反應條件控制不當會發生不完全匹配的雜交以及DNA鏈的自融合現象。而後一現象將在DNA鏈的線性結構上產生髮卡式結構，這種異形結構在電泳過程中影響DNA的行走速度。此外，在生物操作中，注射與震盪會使DNA鏈斷裂；操作試管可以使DNA鏈黏在試管壁、抽筒和吸管上，使計算結果丟失等等。

另一類是生物操作精度與數學解題精度之間的匹配問題。例如抽取操作是所有求解過程中必需的，無論甚麼抽取結果鏈的方法均有一定誤差，而操作引起的誤差，按當前的技術，能達99%的成功概率已經很好，如果計算過程需600個提取，則最終提取出真實結果的概率只有0.002。最後要提出的問題是有關規模與反應的濃度問題^②。分析了阿德勒曼的實驗，結果表明，對於稀疏的有向圖，溶液中結果鏈的濃度會指數級地下降。似乎可以通過聚合酶鏈式反應(PCR)放大中間產物的濃度，但事實上這樣會導致另一個問題：消耗大量的DNA分子，以至完全不可實現。

要解決上述問題，以及許多其他通往實用的DNA電腦的障礙，看上去應當從三方去研究。首先是在生化技術與設備方面，比如威斯康辛大學麥迪遜校區已經實現在固體表面的DNA計算方式^③，然而必須面對的是工作表面的反應大

大減少了DNA運算中的存貯優勢；又好比利用微型機器人來操作生化實驗，通過微型化及規範化可以在相當程度上解決問題；新近在「基因開關」與「基因種」的研究成果帶動DNA計算領域使反應提高可控性，還有像普林斯頓的利普頓提倡以RNA代替DNA用於計算^④，等等。第二個方面是立足於計算模型與演算法的設計，前面介紹的自裝配模型，對於不同的DNA分子結構作為計算資源不僅對應於不同的形式語言，而且這些結構在反應中對穩定性、靈敏與誤差影響方面均有很明顯的影響。因此通過研究自裝配模型的構造，模型的複雜度與施於模型上的操作程式(演算法)，可能是解決這一問題十分有效而經濟的方法。第三個方面是界定DNA電腦所適宜解決的問題類。雖然從理論上講，DNA電腦是完備的，但是一個本質並行的計算與一個本質上串列的計算應當是各有利弊的。已知人腦的並行處理機制在識別一個人的面孔時比現代最快的電腦還快，但是計算一個10位數的50次方就遠遠不如電子電腦。是否未來的世界是DNA電腦與電子電腦或其他類型的電腦分區專利的呢？如果是這樣，界定DNA計算的領域將使實用化研究的進程大大加快。

雖然DNA電腦目前還只是理論設想，現階段的實驗還遠未給出其計算裝置的可行圖景。但許多專家認為，DNA電腦研究的水平猶如電子電腦在30、40年代的情形，在這一領域投入精力將會使科學界甚至經濟領域產生十分誘人的前景。當然，在新型電腦的開發中還有其他一些設想和模式，比較突出的有三種：光學電腦，利用光子流攜帶資料和鐳射光束的彼此穿透來製造微處理器，已經製造了樣機模型；分子電腦，以單個分子和電子代替矽電晶體，起到類似邏輯門和電子開關的作用，但目前只停留在理論階段；還有就是量子電腦，以鐳射或無線電電波穿越原子核列陣，改變原子核的自旋方向，通過其自旋方向的變更進行複雜運算。它們都有各自的優勢和困難，與DNA電腦形成互相競爭的局面。

四 DNA數學的潛在意義

此外，科學家的大腦中還存在另一個十分誘人的精靈，它在生物學世紀中將滲透於科學界，這就是DNA數學。當人們十分驚奇地面對生物世界包羅萬象的多樣性，生物體為適應環境而發育出準確有效的應對環境的群體特徵，以及生物遺傳的嚴格規律性之時，自然會問：在这一切神秘現象的背後，有甚麼力量在支配着？自從人們發現DNA自己不僅可以計算出整個生命世界，而且還可以被人類用於計算人類數學之後，有理由相信，這一切取決於更深層次的電腦制——生物的數學理論，即生物自身的數學。換句話說，生物本身是十分複雜的計算的產物，而這種計算本身又是DNA自身包含的計算規則——DNA數學決定的。這種數學與人類數學大相逕庭，不僅由於它完全不理睬我們賴以計算的「加」「減」，而用「剪」「接」進行運算，而且本質上平行計算的邏輯也完全不顧串列邏輯的因果真理性。因此，一些學者認為，1994年11月是世紀的一門新學科誕生的時間——現代生物數學的誕生^⑤。也許有一天數學家不得不在DNA計算

面前放下自己繁複和艱澀的理論的自豪感，而向「DNA們」學習「剪接數學」。因此DNA計算的意義將不只是產生有巨大效益的DNA電腦，而且還開關了一個新的科學領域。

DNA全序列的資訊破譯與DNA計算有兩個共同的特徵：都以DNA為物質基礎，又都以計算（數學）為基本的理論工具。這顯然不是巧合，因為它們都揭示着同一個新的自然觀：生物是計算的產物，計算是生物的本能。也許有一天人們認識到薛定諤 (Erwin Schrodinger) 對人類最偉大的貢獻並非量子力學；這位給出波動方程的、精通數學的物理學家，在*What Is Life*一書中（沃森 [James Watson] 和克里克 [Francis Crick] 都由於這本小書的啟發而去發現DNA的雙螺旋結構的）啟發人類去發現DNA結構和導致DNA數學的發現，才是「永恆」造福於人類的貢獻。

註釋

①② Leonard M. Adleman, "Molecular Computation of Solutions to Combinatorial Problems", *Science* 266, 1021-24 (November 1994).

③⑨⑩ G. Paun, G. Rozenberg and A. Salomaa, *DNA Computing* (New York: Springer, 1998).

④ 參見註①及D. K. Gifford, "On the Path to Computation with DNA", *Science* 266, 993-94 (November 1994).

⑤ Richard J. Lipton, "DNA Solution of Hard Computational Problems", *Science* 268, 49-66 (April 1995).

⑥ Dan Boneh, Richard J. Lipton and Chris Dunworth, "Breaking DES Using a Molecular Computer", in *Proceeding of DIMACS Workshop on DNA Computing* (1995).

⑦⑧ Rudolf Freund, Lila Kari and Gheorghe Paun, *DNA Computing Based on Splicing: The Existence of Universal Computers* (Technical Report 185-2/FR-2/95, Technical University Wien, 1995).

⑩ Dan Boneh et al., "On the Computational Power of DNA", *Discrete Applied Mathematics* 71 (1996): 79-94.

⑫ Stuart A. Kurtz, Stephen R. Mahaney, James S. Royer and James Simon, "Active Transport in Biological Computing", in *Proceedings of the 2nd DIMACS Workshop on DNA-based Computers* (1996): 111-21.

⑬ Qinghua Liu et al., "DNA Computing on Surfaces", *Nature* 403, 175-79 (13 January 2000).

⑭ Dirk Faulhammer, Anthony R. Cukras, Richard J. Lipton, and Laura F. Landweber, "Molecular Computation: RNA Solutions to Chess Problems", *Proceedings of the National Academy of Sciences of the United States of America* 97 (2000): 1385-89.

⑮ Lila Kari, "DNA Computing: Arrival of Biological Mathematics", *The Mathematical Intelligencer* 19, no. 2 (1997): 9-22.

孟大志 北京工業大學應用數理學院教授

陸 薇 北京工業大學碩士研究生