

蛋白質中殘基的相互作用

中央大學物理系暨生命科學系 李弘謙

蛋白質 (protein) 是自然界最微小的機器。一個普通大小的蛋白質直徑約五奈米 (5×10^{-9} m)，大約是一根頭髮直徑的萬分之一。它也是自然界最神奇的機器，因為這部機器從組裝、尋找受體、執行繁雜的任務、到完成任務後機器的拆散，都是完全自動的。我們每一個人的體內每天都會製造及拆散成千上萬個不同的這種機器。

蛋白質是一切生物藉以表現生命的基本單元。任何一個生命體的繁衍、新陳代謝、結構與運動、乃至於蛋白質本身的製造，都是經由許多不同蛋白質的配合運作而得以順利的執行。每一個蛋白質都由二十種氨基酸 (amino acid) 依特訂的秩序串聯而成。不同的蛋白質有不同長短的氨基酸序列。人體中有上百萬種不同的蛋白質。而比我們原始許多，肉眼無法辨認的微小的美麗線蟲 (nematode) 也有近兩萬種不同的蛋白質。許多人類的蛋白質與相對應的線蟲蛋白質兩者氨基酸序列與三維空間結構都很相似，而功能也相同。這是因為它們有共同的祖先。有這種關係的蛋白質被稱為同源 (homologous)。現有的大量數據顯示大多數的蛋白質都有同源體 (homologs) 分佈在許多不同的生物體中。這個現象是物種演化論的最好的證明。

正因為蛋白質是一切生物藉以表現生命的基本單元，它已經成為現代生醫與藥物學的研究焦點。而瞭解蛋白質的結構與功能也是現代跨領域科學研究的一大重要課題。

蛋白質為了要執行特定的生物功能，必須有特定的空間結構。這就如一個工具或一部機器必須有符合它功能的形狀和結構。如果有某種原因令生物體內一個蛋白質的空間結構產生錯誤而使該蛋白質失去它應有的功能，其後果小則使生物體的運作輕微失調，大則致生物體於死亡。狂牛病就是後者一個近來引起大眾注意的例子。蛋白質三維空間結構的形成有著一個複雜、有趣、奇妙、令人歎為觀止的程序。簡單的說，以人類為例，我們的基因體 (genome) 所含有的最重要的資訊，就是由鹼

基 (nucleotide) 序列編成的基因 (gene)。大部分的這些基因所載的是一個人正常運作所需的所有的蛋白質的編碼 (小部分是核糖核酸 (RNA) 的編碼)，每一個基因載著一個或多個蛋白質的編碼。這種編碼記載著製造蛋白質所需的一切資訊。人類的基因體分別儲藏在二十三個染色體內。而人體內所有的每一個細胞裡都有一套完整的染色體。這就是說每一個細胞裡都藏有一個人正常運作所需的所有的蛋白質的編碼。理論上這似意味著人類的任何一種蛋白質都有可能人體內任何一個細胞被製造。事實上，人體內 (其他生物亦同) 不同部位的細胞無論形狀或功能多已特殊化，而功能特殊化的細胞也只會製造與該功能有關的蛋白質。

蛋白質在細胞裡的製造分兩大步驟：轉錄 (transcription) 和轉譯 (translation)。在有核細胞裡轉錄在細胞核中發生，翻譯在細胞核外的細胞質中發生。轉錄基本上是簡單的把對應於要被製造的蛋白質的編碼從基因體抄錄下來，複製成一段方便於運輸和操作的含載編碼的信使核糖核酸 (mRNA)。轉譯則是將信使核糖核酸上所載的編碼譯出，並依照譯出的指示利用細胞質內已現存的氨基酸串聯合成一個對應於要被製造的蛋白質的氨基酸序列。轉譯是一個複雜的程序，由數種很大的核糖體核糖核酸 (rRNA)，數十種轉移核糖核酸 (tRNA) 和數十種蛋白質參與執行。生物為甚麼不直接就在基因體上的基因做轉譯，而採用了多一道手續的途徑，用信使核糖核酸做仲介呢？我們相信這是生物經過長時間演化出的一個保護基因體上寶貴的遺傳信息的策略，為著使基因體免於在繁雜的翻轉譯過程中因偶發的事件而受損。這個二段式的運作正是一個有經驗的電腦使用者絕不悖離的運作模式：把所有的程式原稿存在永久儲存的硬碟上，要執行某個軟體程式之前，先從硬碟上抄錄一份程式的暫時備份，然後才執行這個備份。這樣，執行時如果出狀況——這經常會發生——就只能損傷備份而不會損害到硬碟上原稿的真實與完整性。

我們很容易看出這種運作模式中的重要步驟與生物運作模式中的重要步驟有一一對應的關係：永久儲存硬碟對應基因體、抄錄對應轉錄、暫時備份對應信使核糖核酸、執行對應轉譯。

基因經過轉譯的產物是一組還沒有生物功能的串聯的氨基酸。這時，或當蛋白質的特殊生物功能不被強調時，這種氨基酸串列常被通稱為多肽鍊（polypeptide），或簡稱肽鏈（peptide），組成肽鏈的氨基酸則通稱為殘基（residues）。生物體中的肽鏈大小不一，可小至十數或大至數千殘基。初生的肽鏈沒有一定的形狀，但是在細胞質裡肽鏈會很快的（視肽鏈長短而定，約在百分之一秒至十秒之間）、自動的折疊成一個有特定形狀的結構，而成為一個有特定功能的蛋白質。蛋白質的折疊過程中有時會受其它蛋白質的幫助，但是蛋白質的空間結構基本上由它的氨基酸序列決定。也就是說蛋白質的空間結構是由殘基與水分子（因為細胞質的成分絕大部分是水）以及殘基與殘基之間的相互作用決定。目前蛋白質的結構必需經過冗長、耗錢又耗力的實驗（主要用 X 光繞射法與核磁共振法）才能確定。如何從肽鏈的氨基酸序列快速預測蛋白質的三維空間結構是生物、病理與藥物等學界共有的一個極待解決的問題。這是一個純物理（或化學）計算但難度極高的問題。今天最先進的超電腦要花數年才能算出一個簡單的蛋白質的結構。為了解決這個問題，IBM 公司目前正在全力設計一個專門算蛋白質結構的超級電腦組 [1]。

氨基酸是一種由碳、氫、氧、氮及硫組成的含十至三十個原子的大分子。兩個氨基酸（或殘基）之間的相互作用理論上可以用電磁理論算出。實際上這個問題因為牽涉的原子太多（水也扮演重要的角色，見下）而非常複雜，我們目前對它只有粗略的瞭解。退而求其次，人們依據一個簡單的想法用現象學的方法從已知結構的蛋白質資料庫中求出殘基之間的有效相互作用。這個想法是：如果兩個殘基之間的相互吸引力愈大，則在蛋白質中它們為近鄰的機率也愈大。當然兩個殘基在蛋白質中是否為近鄰要受許多其它因素的影響，尤其是殘基序列的影響，但是這些因素在不同的蛋白質中是雜亂、隨機的，所以在許多蛋白質的總效應中會相互抵銷。因此如果有很多蛋白質的結構數據，可以求出殘基之間的相對相互吸引力。因為有二十種氨基酸，這類計算的結果是一個 20

乘 20 的對稱矩陣，每一個矩陣元就代表一對殘基的有效相互作用能量。一般認為這一類矩陣中 Miyazawa 和 Jernigan [2] 所求得的最具代表性。一個一般的 20 乘 20 的對稱矩陣原應有 210 個獨立量，但是湯超等 [3] 發現 Miyazawa 和 Jernigan 的矩陣，減去平均值後，其實只有 22 個獨立量。這結果讓我們可以把 Miyazawa 和 Jernigan 的矩陣重組成一個較簡單的形式，使我們認識到 22 個獨立量中的 20 個代表二十種氨基酸各在已折疊蛋白質中的（對所有已知結構蛋白質取的）相對平均位能，一個獨立量代表位能強度，最後一個獨立量代表殘基之間相互作用的偶合強度。同時我們也知道位能比殘基之間相互作用能大一個數量級。這個形式說明決定蛋白質粗粒結構的主要因素不是殘基之間的相互作用，而是殘基在已折疊蛋白質中所感受的平均位勢場 [4,5]。

有趣的是，以上所述的 22 個獨立量可以從物理的出發點瞭解。如果用水的表面張力對殘基所致的排斥、水分子和各殘基的電偶極、各殘基與水分子之間的電偶極相互作用、水分子和平均殘基的大小、親水及厭水兩類殘基在已折疊蛋白質中的掩埋度等因素作概算，我們能大約的導出與殘基在已折疊蛋白質中所感受的平均位勢場有關的 21 個獨立量，最後一個獨立量則可由殘基與殘基之間的電偶極相互作用大約導出 [4]。這說明相當簡單的物理已能為殘基在已折疊蛋白質中的平均作用能在大分子層次給予一個不錯的描述。當然，我們以上談的是一個非常粗粒化的描述。大分子層次的相互作用是不足以描述任何一個特定蛋白質的結構細節的，這種結構細節取決於原子與原子之間在多體環境中的相互作用。目前我們對這些複雜的相互作用還沒有很可靠的瞭解。

即便如此，人們仍然嘗試著用近似的原子相互作用較細粒化的模擬蛋白質的結構。如上所說，這類模擬的計算量很大，非極大量的平行計算模式無法勝任。

一個很聰明的作法是將模擬的電腦程式寫成螢幕保護程式，並讓大量的個人電腦在主人不用它時靜靜的跑模擬蛋白質結構的計算 [6]。我們目前正在規劃在台灣執行一個類似的計畫。

參考文獻

- [1] "The IBM Blue Gene team", IBM Journal,

- 40**, 310(2001); 參閱網頁
http://www.genomicglossaries.com/content/protein_structure_glossary.asp
及
<http://www.research.ibm.com/bluegene/>。
- [2] S. Miyazawa and R. L. Jernigan, *Macromolecules* **18**, 534 (1985); *J. Mol. Biol.* **256**, 623 (1996).
- [3] H. Li, C. Tang and N.S. Wingreen, *Phys. Rev. Lett.* **79**, 765 (1997).
- [4] Z.H. Wang and H.C. Lee, *Phys. Rev. Letts.* **83** (2000) 574.
- [5] O. Keskin, I. Bahar, A.Y. Badretdinov, O.B. Ptitsyn and R.L. Jernigan, *Protein Science* **7**, 2578 (1998).
- [6] V. Pande, “The FOLDING@HOME Project”, *Science* **290**, 1903(2000).