

談蛋白質折疊與氨基酸序列

王自豪*§ 林誠謙*# 李弘謙§

*中央研究院計算中心 #中央研究院物理研究所 §國立中央大學物理系

§ E-mail: claude@gate.sinica.edu.tw

*#E-mail: sclin@sinica.edu.tw

§E-mail: hcllee@phy.ncu.edu.tw

摘要

蛋白質(Protein)是一切生物藉以表現生命功能的基本單元。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形成，都是經由許多不同蛋白質精密的配合運作而得以順利的進行。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。不同的蛋白質則由不同長短的氨基酸序列組成。蛋白質執行各種特定的生物功能則完全仰賴其特定的三度空間結構。Miyazawa 和 Jernigan 兩人在 1985 和 1996 年間，從蛋白質資料庫中統計分析所有已知結構的蛋白質內部不同種類氨基酸間相互接觸的數目，再利用統計熱力學的波茲曼原理導出任兩種氨基酸間，平均在所有蛋白質結構中相互接觸的約略能量。利用這樣的數據資料，我們可以定性及定量的解釋，驅動蛋白質氨基酸序列折疊形成特定三度空間結構的作用力，主要是來自於氨基酸旁鏈(side chain)分子和水分子間的表面張力及其電偶極(dipole)間電磁力的交互作用，和氨基酸旁鏈分子間電偶極的電磁力交互作用。這個新的物理詮釋或許可以幫忙應用在改善氨基酸序列早期快速折疊的分子動學模擬。

蛋白質-Protein 是取用自希臘字 *protos*，意指“第一”的意思。它是一切生物藉以表現生命的最重要基本單元，它可以算是自然界最微小的自動機器。一個普通大小的蛋白質直徑約五奈米($5 \times 10^{-9} \text{m}$)，大約是一顆蛋直徑的千萬分之一。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間精密的配合運作而得以順利的進行。為了要執行特定的生物功能，蛋白質必須擁有特定的三度空間結構。這好比是一個工具或一部機器必須有符合它功能的形狀和

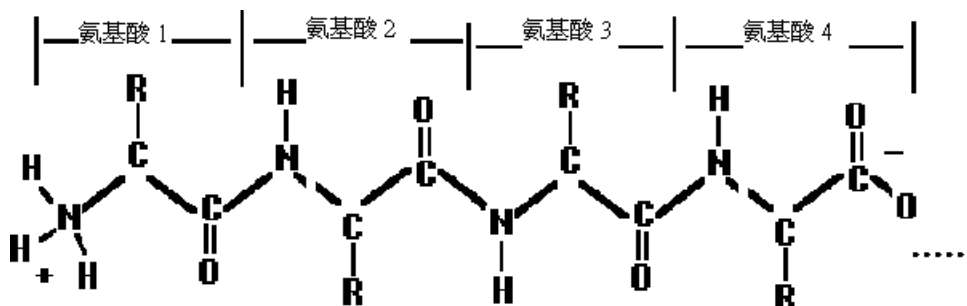
結構才具有用途。如果有某種原因令生物體內一個蛋白質的空間結構產生錯誤而使該蛋白質失去它應有的功能，其後果小則使生物體的運作輕微失調，大則導致生物體死亡。狂牛病的起因就是上述的一個典型例子。蛋白質三度空間結構的形成，背後隱藏著一個極端複雜、令人歎為觀止的過程。簡單的說，以人類為例，我們的基因體(genome)所含有的最重要的資訊，就是由鹼基(nucleotide)序列所編成的基因(gene)。所有基因裏所載的鹼基序列是一個人正常運作所需所有的蛋白質的編碼，每一個基因

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條染色體內。而人體內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蛋白質的編碼。理論上這似乎意味著人類的任何一種蛋白質都有可能在人體內任何一個細胞被製造。事實上，人體內（許多其他生物亦同）大部份的細胞在胚胎發育時都已經過特化的過程，而功能特化的細胞也只會製造與該功能有關的特定種類蛋白質。蛋白質在細胞裡經過所謂轉錄(transcription)和翻譯(translation)的過程而被合成。轉錄基本上是簡單的把對應於要被製造

的蛋白質的鹼基編碼從基因上抄錄下來，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)。翻譯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，串聯成一個對應於要被製造的蛋白質的氨基酸序列。翻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。

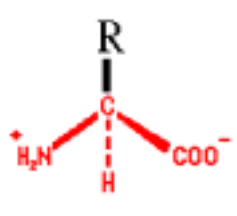
基因經過翻譯的產物是一組還沒有生物功能的氨基酸串列。這樣的串列是由二十種氨基酸依特訂的秩序串聯而成的，通稱為蛋白質的一級結構(Primary Structure, 圖一)。

圖一



每一個氨基酸都有一個中心 α -碳原子(C), C (COOH)和一個被稱為旁鏈(side chain)的 R 基團相連(圖二)。

圖二

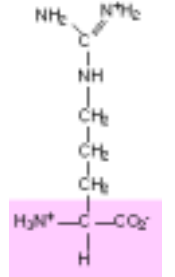
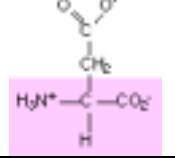
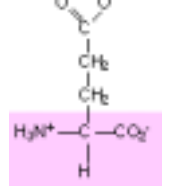


二十種氨基酸的差別在於與 C 相連的旁鏈基團 R 的不同。根據氨基酸的名稱、所帶電荷、和化學性質，二十種氨基酸可以分類如下表一：

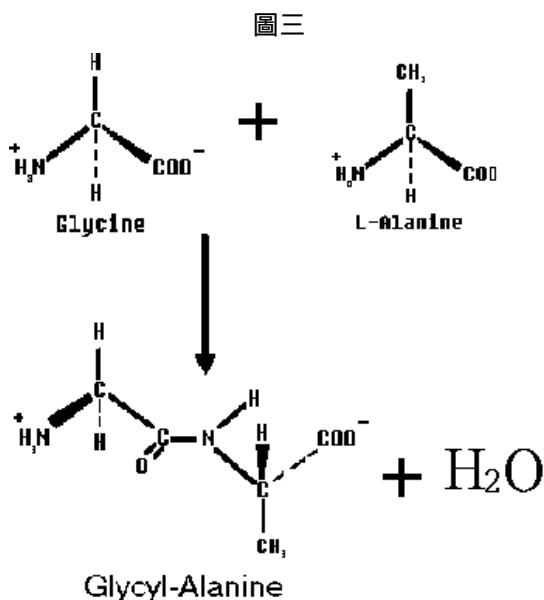
表一

分類	氨基酸名稱	英文縮寫 (符號)	性質	化學結構
	甘氨酸 glycine	Gly (G)	親水	
脂 肪 族	丙氨酸 alanine	Ala (A)	厭水	
	纈氨酸 valine	Val (V)	厭水	
	亮氨酸 leucine	Leu (L)	厭水	
	異亮氨酸 isoleucine	Ile (I)	厭水	
	脯氨酸 proline	Pro (P)	厭水	
	半胱氨酸 cysteine	Cys (C)	親水	
	甲硫氨酸 methionine	Met (M)	厭水	
	芳 香 族	組氨酸 histidine	His (H)	親水

	苯丙氨酸 phenylalanine	Phe (F)	厭水	
芳香族	酪氨酸 tyrosine	Tyr (Y)	親水	
	色氨酸 tryptophan	Trp (W)	厭水	
極性	天冬醯胺 asparagine	Asn (N)	親水	
	谷氨醯胺 glutamine	Gln (Q)	親水	
	絲氨酸 serine	Ser (S)	親水	
	蘇氨酸 threonine	Thr (T)	親水	
帶電荷	賴氨酸 lysine	Lys (K)	帶正電荷	

	精氨酸 arginine	Arg (R)	帶正電荷	
	天冬氨酸 aspartate	Asp (D)	帶負電荷	
帶電荷	谷氨酸 glutamate	Glu (E)	帶負電荷	

氨基酸串列的形成是經由一個氨基酸的酸基和另一氨基酸的氨基縮合脫去一個水分子而形成一肽鍵(圖三, 以氨基酸-Gly 和氨基酸-Ala 為例說明),



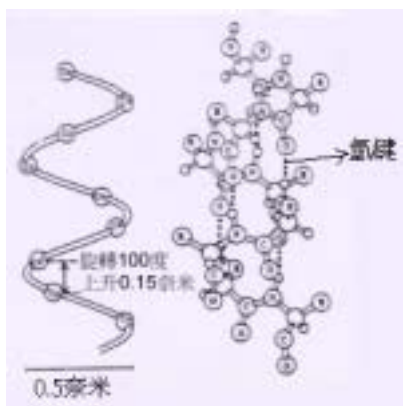
多個氨基酸, 通常由幾十到一百個, 由肽鍵相

連形成多肽鏈(polypeptide), 或簡稱肽鏈(peptide)。

多肽鏈中由一個氨基、一個 C 碳原子以及一個酸基的重覆單位構成一骨架(backbone) 骨架中從氨基的氮原子(N)到酸基的碳原子(C)的一個單位稱為殘基(residue)。多肽鏈的兩端構造並不同, 因此它是有方向性的, 按慣例, 氨基端是起始端, 向 C 端延伸。生物體中的肽鏈大小不一, 可包含小至數十或大至數千個氨基酸。初生的肽鏈沒有一定的形狀, 但是在細胞質含水的環境裡肽鏈會很快的(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三度空間結構, 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。由於蛋白質是在水溶液的環境中折疊並行使其生物功能, 為減小其厭水旁鏈與水介質的相互作用, 蛋白質形成空間結構必須遵守一個重要原則: 將厭水旁鏈埋入分子內部而將親水旁鏈暴露在表面與水接觸, 由於蛋白質骨架上每一個氨基酸都有一個氫鍵的給體(NH)和一個氫鍵的

受體(C=O)，且骨架是高度親水的，在將厭水旁鏈埋入分子內部時，必將親水的骨架也埋入內部。為解決這個矛盾，大自然在演化的過程中，選擇了一個非常巧妙的策略，就是在分子內形成所謂的二級結構(Secondary Structure)，其中最重要的有 α -螺旋 (α -helix)和 β -薄片(β -sheet)。

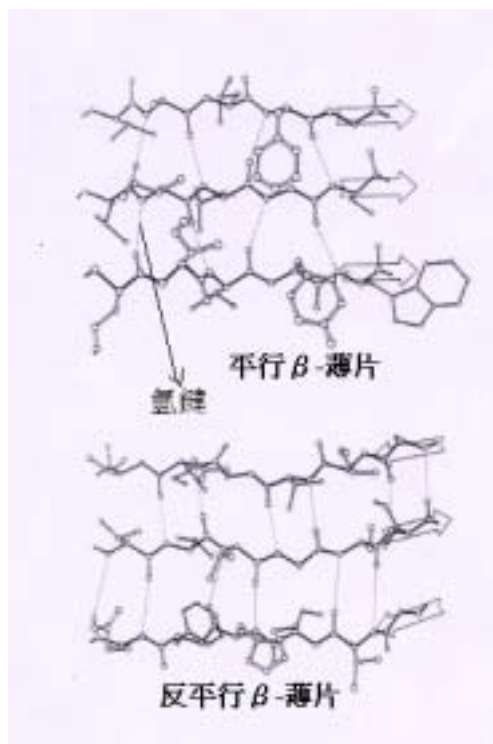
圖四



α -螺旋(圖四)是多汰鏈的一種螺旋形的空間結構，汰鏈繞著一根中軸旋轉，每一個氨基酸殘基都沿著軸旋轉 100° ，向上平移約 0.15 奈米。螺旋每上升一圈包含 3.6 個氨基酸，螺距約為 0.54 奈米。

α -螺旋體中氨基酸殘基的側鏈伸向外側，相鄰的螺圈之間形成一氫鍵，是由氨基酸上面的氫和在它後面約第四個氨基酸殘基的酸基氧之間形成的，其取向幾乎和中軸平行，上述的螺旋結構是自然界中一個非常穩定的結構。

圖五



β -薄片(圖五)分為平行和反平行兩種，構成 β -薄片的兩段汰鏈走向(即從氨基到酸基排列方向)相同的為平行的 β -薄片，而方向一正一反則為反平行的 β -薄片。相鄰的兩段汰鏈之間相互形成氫鍵。特點是殘基在 β -薄片的兩側，指向呈正反交錯。整個多汰鏈的各部分所折疊成的二級結構彼此再經由汰鏈的轉折(turns)及迴圈(loops)連結而形成一個最穩定的三度空間結構就是具有特定生物功能的蛋白質結構，亦稱作三級結構(Tertiary Structure)。如圖六所示為兩個典型的蛋白質三級結構。

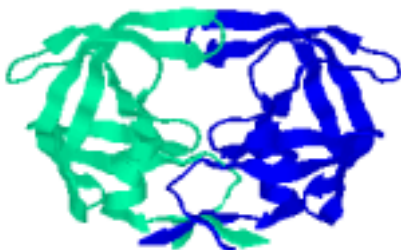
圖六



紅色部份為 α -螺旋, 綠色部份為 β -薄片, 黑色部份為連接 α -螺旋或 β -薄片的轉折(turns)或迴圈(loops)結構。

但有一些較複雜的大型複合蛋白質是由一條以上的多汰鏈所組成, 因此一個以上的三級結構會再形成一更穩定且功能性的複合結構, 這樣的結構則被稱為四級結構(Quaternary Structure)。如圖七所示即為一種愛滋病毒蛋白雙聚體的四級結構。

圖七



一種愛滋病毒蛋白雙聚體的四級結構

目前蛋白質的結構必需經過冗長的實驗程序(主要用 X 光繞射法與核磁共振法)才能確定。如何從汰鏈的氨基酸序列快速預測蛋白質的三度空間結構一直是生物、病理與藥物等學界共同急待解決的問題。這是一個純物理(或化學)計算但難度極高的問題。就連今天最先進的超級電腦也要耗費數年的光陰才能完全解出一個由幾十個氨基酸所組成的蛋白質的結構。人體中估計約有十萬種不同的蛋

白質。而比我們原始許多, 肉眼無法辨認的微小線蟲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。這種序列、結構及功能都相近的蛋白質被稱為同源蛋白質, 也就是說它們在生命演化上有共同的起源。現有的大量數據即顯示大多數的蛋白質都有其同源體分佈在許多不同的生命體中。這個現象是物種演化論的最好的證明。關於蛋白質種種方面的研究一直是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則是現代跨領域科學研究的一大重要課題。

使 Anfinsen 得到諾貝爾獎的一個重大研究工作^[1]明白告訴我們, 一般蛋白質的結構完全可以由其組成的氨基酸序列所決定。也就是說蛋白質的空間結構是由氨基酸與水分子以及氨基酸與氨基酸之間的相互作用決定。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(水分子也大量的牽涉其中)而非常複雜, 我們目前對它只有粗略的瞭解。物理學家喜歡化約式的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。(統計熱力學中的玻茲曼原理指出, 在熱力平衡態下, 系統某一組態出現的機率正比於

$$e^{\left(\frac{-\text{組態能量}}{\text{玻茲曼常數} \times \text{溫度}}\right)}$$

), 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，這類計算的結果自然是一個 20 乘 20 的對稱矩陣，每一個矩陣元素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 的對稱矩陣原本應該有 210 個獨立量，但是國際上在 NEC 研究中心，湯超的研究團隊^[3] 發現 Miyazawa 和 Jernigan 的矩陣，減去平均值後，經過線性代數的分析，其實只剩下 22 個獨立量。這結果使我們可以把 Miyazawa 和 Jernigan 的矩陣重寫成一個較簡單的形式，我們進一步理解到這 22 個獨立量中的 20 個量恰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所取的）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]。

以上所述的 22 個獨立量皆可以從物理的出發點瞭解。如果用水的表面張力對殘基所產生的排斥、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包埋度等因素作概算，我們能約略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]。我們的結果已經能成功的以百分

之九十的準確度解釋並吻合 Miyazawa 和 Jernigan 從蛋白質資料庫所統計出的矩陣數據。

當然，我們以上談的仍是一個非常粗粒化約的描述。在整個殘基之大小尺度上的相互作用，是不足以描述任何一個特定蛋白質結構細節的，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則還沒有很可靠的瞭解。即便如此，人們仍然能嘗試運用此近似的粗粒化約模型，在整個殘基之大小尺度上，花費較短的計算時間，模擬蛋白質早期快速折疊的過程，提供人們更多對蛋白質折疊過程的瞭解。

參考文獻

- [1] C. Anfinsen, *Science* **181** (1973) 223.
- [2] S. Miyazawa and R.L. Jernigan, *Macromolecules* **18** (1985) 534; *J. Mol. Biol.* **256** (1996) 623.
- [3] H. Li, C. Tang and N.S. Wingreen, *Phys. Rev. Lett.* **79** (1997) 765.
- [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** (1998) 2578.