

用平行電腦及解析方法研究巨分子

胡進錕¹、林財鈺²

中央研究院物理所

¹e-mail: huck@phys.sinica.edu.tw

²e-mail: lincy@phys.sinica.edu.tw

摘要

本文所稱之巨分子包括同型聚合物(homopolymer)和異型聚合物,後者包括蛋白質、DNA 和 RNA 等生物巨分子。巨分子之結構及相變與許多重要的學術及實用問題有關。本文簡單介紹這方面研究的近況以及筆者與合作者所做的一些工作。

蛋白質(protein)是組成生物極重要的巨分子,由 20 種胺基酸(amino acid)靠化學鍵連接成一維的系列。這一維的系列在生物體內產生後,會在很短的時間(1 μ s-1s)摺疊成立體結構。而蛋白質的生理功能和它的立體結構有極密切的關係,因此許多科學家想各種辦法決定蛋白質的立體結構。

用實驗方法決定蛋白質立體結構的工具主要有 X 射線和 NMR,這些設備很昂貴,且取得實驗數據後,用數值分析決定蛋白質結構通常也要花很長時間。如果能夠純粹用電腦計算就算出蛋白質的立體結構,那將是很有意義的事,但這是不是可能呢?

根據諾貝爾獎得主 Anfinsen^[1]的研究,不是特別長的蛋白質,一維系列的資訊就能決定蛋白質的立體結構,而室溫時蛋白質的立體結構就是它在基態或接近基態時的結構,因此許多科學家也想各種辦法由蛋白質一維的資訊計算它的立體結構。常用的計算方法包括分子動力學(molecular dynamics,

MD)模擬法和蒙地卡羅(Monte Carlo, MC)模擬法。用 MD 或 MC 計算蛋白質低溫(室溫)狀態常碰到的困難,就是蛋白質容易掉入暫穩狀態而不容易再跳出來找到更低的能量狀態。為了克服這個困難, Hansmann 和 Okamoto^[2] 首先將多重正則 MC(multicanonical MC, MMC)用於模擬由 5 個胺基酸組成的 Met-enkephalin 而得到不錯的結果。MMC 的特點就是讓分子在能量空間隨機行走,因此它比較容易走到低能量的狀態。另一種容易找到低能量狀態的方法就是平行調整法(parallel tempering)^[3],即同時在 M 部電腦模擬 M 種不同溫度蛋白質之動態行為,過一段時間,再以某種機率決定不同電腦上之蛋白質是否要交換溫度,然後再重複前述動作,此一方法的優點就是容易在平行電腦上運算。

蛋白質的立體結構和它與周圍水的作用有很密切的關係,處理蛋白質和水的作用也有各種辦法。

用 MD 研究蛋白質的科學家，通常將大量的水分子 (如 3000 個) 和蛋白質放在一起用 MD 模擬它們間的互動情況，另一種方法就是將水當作是連續體，即將蛋白質放在水的連續體內，模擬蛋白質在水作用下的行為；這種方法所需的計算機時間比 MD 要少很多，但關鍵的問題就是要找出合適的蛋白質與水作用的模型。

以 D 模擬蛋白質，極有名的例子就是加州大學舊金山校區 Duan 和 Kollman^[4] 用 Cray T3E (包含 256 個 CPU) 計算 villin headpiece subdomain (包含 36 個胺基酸，簡稱 HP-36) 在 3000 個水分子中摺疊的情況，他們計算 HP-36 由無秩序狀態經過大約 1 μ s 後達到接近 HP-36 在室溫 (由實驗決定) 的立體結構，但兩者仍有明顯的差異。

另一種蛋白質的理論與數值研究是以簡化的晶格模型模擬蛋白質在空間的結構。代表性的研究是李浩、Helling、湯超和 Wingreen^[5] 利用蛋白質中之胺基酸可分類成疏水性 (H) 和親水性 (P) 的 HP 表示式^[6] 在數個小正方形和 3x3x3 立方晶格上的研究。經由這個研究，他們提出可設計性 (designability) 的概念以了解什麼樣的胺基酸序列容易摺疊成蛋白質。

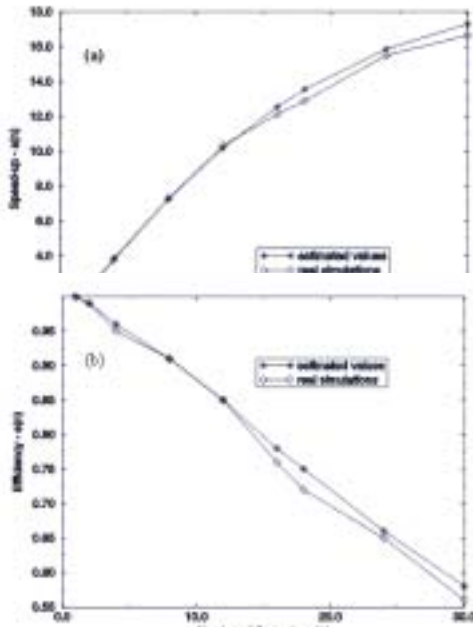
RNA 則是另一種重要的生物巨分子，由 A (adenine)、U (uracil)、C (cytosin) 和 G (guanine) 四種核苷酸的鹼基組成，其中 A 和 U 及 C 和 G 可由氫鍵結合。若 N 個鹼基組成之 RNA 一維系列由 1, 2, …… 至 N 依序編號，其中 i 號和 j 號由氫鍵結合，k 號和 l 號也由氫鍵結合， $i < j$ ， $k < l$ ， $i < k$ ，且四個整數都不相同。兩對鹼基有下列三種關係 (1) $i < j < k < l$ ，即 i 和 j 及 k 和 l 形成的結 (knot) 並列，(2) $i < k < l < j$ 即 i 和 j 形成的結中間包含 k 和 l 形

成的結，(3) $i < k < j < l$ ，即 i 和 j 形成的結與 k 和 l 形成的結交錯，稱為交錯結 (pseudo knot)。目前通行的 RNA 理論^[7]，為了計算上的方便，將交錯結完全忽略不計，但在實際 RNA 的結構中，交錯結的確存在，且在 RNA 之結構及功能扮演重要角色^[8]。最近 Zee 和 Orland 以 NXN 矩陣場論表達 RNA 摺疊問題^[9]，此理論在 $1/N^2$ 的修正項有考慮交錯結的貢獻。

我們曾以 MC、重整化群及解析方法研究自旋玻璃、Ising 模型、Potts 模型、展透模型、沙堆模型等相變或自組成臨界性的模型^[10]。近年來則開始與合作者用 MC 及解析方法研究蛋白質、DNA、RNA、及同型聚合物，所得之主要成果計有：

1. 完成可用於計算蛋白質熱力性質與立體結果的套裝軟體 SMMP (Simple Molecular Mechanics for Proteins)^[11]。此套裝軟體用 Fortran 寫成，可用於模擬蛋白質標準幾何模型，蛋白質之內能則包括 ECEPP/2、ECEPP/3 和 FLEX 等模型。蛋白質與水交互作用的能量可由每一原子接觸水的面積及每種原子與水相互作用參數計算出來。此套裝軟體也包括模擬退火 (simulated annealing)，MMC，平行調整 (parallel tempering) 等數值模擬法之計算程式。據筆者所知，SMMP 於 2001 年夏天發表後已有德國 Peter Grassberger 和美國 B. A. Berg 等數值模擬著名教授開始用 SMMP 模擬蛋白質。
2. 發展出計算蛋白質能量之平行算則^[12]，將 MMC 和平行計算結合後所算的 Met-enkephalin 熱力學性質和別人用單一 CPU 電腦算的結果一致。在我們計算能量的平行算則，資料在 CPU 之間傳送的量很低，因此計算效率不至於隨 CPU 數快速下降。我們用中央研究院計算機中心包含

48 個 CPU(每一 node 只有一個 CPU)的 PC Farm 計算 Protein L 能量，測試一週，發現 CPU 數到 30 時，效率仍高達 57%，詳見圖一。

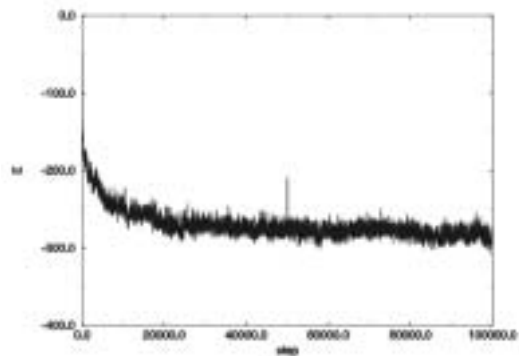


圖一：用平行電腦計算蛋白質 L 能量之測試結果，此平行電腦共有 48 個 Pentium III 500MHz 中央處理器 (CPU)。用 n 個 CPU 完成 34,500 次能量計算所需之時間當 n=1,2,4,8,16,24 和 30 時分別為 180.96, 91,533, 47.75, 24.85, 17.75, 14.917, 14.05, 11.683 和 10.867 分鐘。由此數據可算出 (a) 有效 CPU 數 (Speed-up) 隨 n 之變化；(b) 平行化效率 (Efficiency) 隨 n 之變化。本圖取自參考資料[12]，實線代表實際計算，而虛線代表理論估計值。

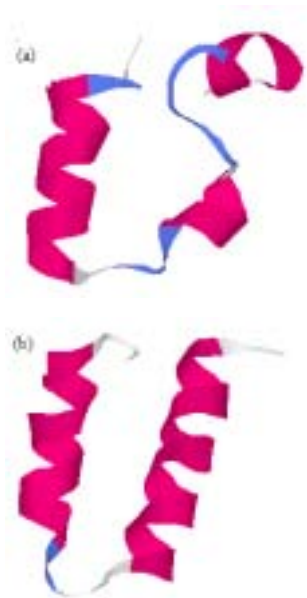
3. 我們用 MC 研究在簡單立方晶格上包含鐵磁性及反鐵磁性最鄰近耦合作用易行(Ising)模型之低溫特性^[13]，反鐵磁性耦合常數所佔的比率為 R。R=0 與 R=0.5 分別相當於純鐵磁性物質與自旋玻璃 (spin glass)。1999 年，Hartman 發現當 R 由 0 增加至 0.5 時，此系統的特性會由鐵磁性過渡到自旋玻璃而臨界點為 $R_c=0.222\pm 0.005$ ^[14]。我們則發現當 R 接近但小於 R_c 時此系統有些基

態頗類似蛋白質，即一個基態可以由許多不同的耦合常數組態所設計。此種類蛋白質的基態在少數耦合常數變異 (即改變正負號) 下比較穩定。

4. 彭仲康 (C. K. Peng) 和他的合作者曾將 DNA 一維系列映射至一維漫步而發現 DNA 有長程相關性。用類似方法 Pande 等人發現一些蛋白質一維系列有持續的相關性而 Irbüch 等人則發現一些蛋白質一維系列有持續的反相關性，即兩研究群所發現之結果不同^[15]。我們用小波 (wavelet) 的分析法研究蛋白質家族胺基酸系列之相關性，發現一些蛋白質家族有持續相關性，另一些蛋白質家族則有持續的反相關性^[16]。
5. 用解析方法計算一個簡化但包含交錯結 RNA 模型二級結構的熔解溫度。我們發現熔解溫度隨溶液離子濃度對數值線性下降，此結果與實驗量測之結果符合^[17]。
6. 提出同型聚合物在多元溶劑發生臨界現象之理論^[18]。此理論可解釋中央研究院物理所杜其永與其合作者所發現的光從聚合物溶液 PEO/NE/MP (PEO 溶於二元溶劑 NE/MP) 散射所產生之異常強散射現象^[19]。



圖二：在溫度 T=250K 下，能量對 MC 步驟之時間序列圖。



圖三：(a) 由 NMR 決定之 HP-36 立體結構圖。(b) 由我們的模擬得到之立體結構圖。

7. 我們與 Hansmann 合作用平行電腦研究 HP-36 的三維結構。在這個研究裏我們完成了一個結合 SMMP 和平行調整法的 MPI 平行化程式。我們用中研院計算機中心的 IBM 平行電腦從隨機組態出發，以 20 個 CPU 模擬 20 個溫度下 HP-36 的平衡性質，透過平行調整法不斷地在各個 CPU 間交換溫度以求得能量的極小值。圖二是溫度 $T=250\text{K}$ 在十萬次 MC 模擬步驟中的能量對時間圖。圖中顯示隨時間演進能量漸漸降低。圖三 (a) 是由 NMR 所決定的 HP-36 立體結構圖，圖三 (b) 是由 20 個溫度，每個溫度十萬次 MC 移動所得之最低能量組態。由圖三，我們可以看出 (a) 和 (b) 左邊結構已有相當的類似性，也許需要更長的模擬時間或修改蛋白質與水作用的模型，才能使得 (b) 右側結構類似 (a) 的右側結構。

目前 SMMP 只能用於模擬蛋白質，以後我們擬以 SMMP 為基礎，發展可用於模擬 DNA 與 RNA 的計算機程式集。

蛋白質晶格模型可能有助於了解形成蛋白質立體結構的機制，但是要由蛋白質的一維系列算出它的立體結構，則有賴於在連續的空間由較接近真實情況的蛋白質全原子模型 (all atoms model)。計算蛋白質的立體結構本文前面介紹第 1, 2 和 7 點的工作代表我們在這一方向所走的一小步。

隨著蛋白質與水交互作用模型、蛋白質內能模型、計算機軟硬體、數值模擬算則，以及數值模擬與實驗數據 (如 NMR 數據) 整合各方向的進展，也許這一小步會變成結構生物學的一大步。

參考文獻：

1. C. B. Anfinsen, *Science* 181, 223 (1973).
2. U. H. E. Hansmann and Y. Okamoto, *J. Comp. Chem.* 14, 1333 (1993).
3. G. J. Geyer, *Stat. Sci.* 7, 437 (1992); M. C. Tesi, E. J. J. van Rensburg, E. Orlandini and S. G. Whittington, *J. Stat. Phys.* 82, 155 圖 (1996); K. Hukushima, K. Nemoto, *Journal of the Physical Society of Japan* 65, 1604 (1996); U. H. E. Hansmann, *Chem. Phys. Lett.* 281, 140 (1997).
4. Y. Duan and P. A. Kollman, *Science* 282, 740 (1998).
5. H. Li, R. Helling, C. Tang, and N. Wingreen, *Science*, 273, 666 (1996); 相關研究可參考 C. T. Shih, et al. *Phys. Rev. Lett.* 84, 386 (2000).
6. K. A. Dill, *Biochemistry* 24, 1501 (1985); K. F. Lau and K. A. Dill, *Macromolecules* 22,

- 3986 (1989).
7. P. G. Higgs, Phys. Rev. Lett. 76, 704 (1996); R. Bundschuh and T. Hwa, Phys. Rev. Lett. 83, 1479 (1999); A. Pagnani, G. Parisi and F. Ricci-Tersenghi, Phys. Rev. Lett. 84, 2026 (2000).
 8. D. P. Giedroc, C. A. Theimer and P. L. London, J. Mol. Biol. 298, 167 (2000).
 9. H. Orland and A. Zee, Nucl. Phys. B 620, 456(2002).
 10. C. Dasgupta, S. K. Ma, and C.-K. Hu, Phys. Rev. B 20, 3837(1979); S. K. Ma, C. Dasgupta, and C.-K. Hu, Phys. Rev. Lett. 43, 1434(1979); C.-K. Hu, Phys. Rev. B1, 46, 6592 (1992) and Phys. Rev. Lett. 69, 2739 (1992); C.-K. Hu, C.-Y. Lin, and J.-A. Chen, Phys. Rev. Lett. 75, 193 and 2786(E) (1995); C.-K. Hu and C.-Y. Lin, Phys. Rev. Lett. 77, 8 (1996); C.-K. Hu, E. V. Ivashkevich, C. Y. Lin, and V. B. Priezzhev, Phys. Rev. Lett. 85, 4048 (2000); N. Sh. Izmailian and C.-K. Hu, Phys. Rev. Lett. 86, 5160 (2001); V. B. Priezzhev, E. V. Ivashkevich, A. M. Povolotsky, and C.-K. Hu, Phys. Rev. Lett. 87, 084301 (2001).
 11. F. Eisenmenger, U. H. E. Hansmann, S. Hayryan, and C.-K. Hu. [SMMP] A modern package for simulation of proteins, Computer Phys. Commu. 138, 192 (2001).
 12. S. Hayrian, C.-K. Hu, S.-Y. Hu, and R.-J. Shang. J. Comp. Chem. 22, 1287 (2001).
 13. C.-Y. Lin, C.-K. Hu, and U.H.E. Hansmann, Phys. Rev. E, 64, 052903 (2001).
 14. A. K. Hartmann, Phys. Rev. B 59, 3617 (1999).
 15. C. K. Peng, et al., Nature (London), 356, 168 (1992); V. S. Pande, A. Yu. Grosberg, and T. Tanaka, Proc. Natl. Acad. Sci. USA 91, 12972 (1994); A. Irback, C. Peterson, and F. Potthast, Proc. Natl. Acad. Sci. USA 93, 9533 (1996).
 16. A. N. Silchenko, E. SH. Mamasakhlisov, and C.-K. Hu. Persistent and anti-persistent correlations in protein families and superfamilies, preprint.
 17. E. SH. Mamasakhlisov, Shura Hayryan, C.-K. Hu, and V. F. Morozov. RNA folding in the presence of counterions, preprint.
 18. Zh. S. Gevorkian and C.-K. Hu. New phase transition in polymer solutions with multicomponent solvent, e-print cond-mat/0201170 and Phys. Rev. Lett, submitted.
 19. K. To, C. A. Kim, H. J. Choi, Physica A 254, 292 (1998).