

利用核磁共振光譜自動化決定蛋白質在水溶液中的三度空間結構

Automated 3D Structure Determination of Proteins in Solution Using NMR Spectra

林意然、Peter Güntert

Yi-Jan Lin, Peter Güntert

藥物設計的主要基礎之一是了解生物大分子結構與功能之間的關係，三度空間結構扮演一個關鍵的角色，因為結構的知識對於了解一種蛋白質的物理、化學及生物的特性是必須的。直到最近，以 NMR 決定蛋白質結構仍然是一項辛苦的工作，對於每一種新的蛋白質結構，光譜學家都得花上數個月到幾年的時間，其實很多耗時的工作，像是光譜的分析，都可藉由自動化的計算系統更有效率地完成⁽¹⁻³⁾。本文將簡介一個自動化 NMR 蛋白質結構計算的方法，CYANA，以及如何利用此方法決定蛋白質的三度空間結構。

Understanding the relationship between structure and function of biological macromolecules is one of the key elements of rational drug design. The three-dimensional structure has a pivotal role, since its knowledge is essential to understand the physical, chemical, and biological properties of a protein. Until recently NMR protein structure determination has remained a laborious undertaking that occupied a trained spectroscopist over several months or few years for each new protein structure. It has been recognized that many of the time-consuming interactive tasks like the process of spectral analysis can be done more efficiently by automated, computational systems. This article will introduce a program of automated NMR structure calculation, CYANA (combined assignment and dynamics algorithm for NMR applications), and how to utilize it to determine three-dimensional structures of proteins.

一、前言

自從四十幾年前第一個蛋白質的核磁共振光譜被發表以來，核磁共振在生物大分子方面的研究及應用，便不斷有新的發展，例如：高磁場核磁共振儀 (圖 1) 的使用、多維 (multiple dimensional) 核磁

共振技術及各種脈衝序列 (pulse sequence) 的發展⁽⁴⁾、以及自動化軟體的應用等，使得 NMR (nuclear magnetic resonance, NMR) 已成為用來研究蛋白質及核酸等大分子的主要技術之一。這些技術的發展，不但提升了儀器的靈敏度 (sensitivity)，也增加了共振訊號的解析度 (resolution)，更大大地加快了



圖 1. 裝置於日本理化學研究所 (RIKEN) 之 900 MHz 核磁共振儀。

決定蛋白質結構的速度。隨著這些技術的不斷研發，可被研究的蛋白質分子量也不斷地提高，一般而言，分子量在 25 kDa 以下的蛋白質，都可以用 NMR 技術決定其結構。

在 1984 年以前，原子層面上結構的資訊，只能由蛋白質的晶體經由 X 光繞射而決定，而 NMR 光譜學技術在決定蛋白質結構方面的應用，已經可以用於解析水溶液中蛋白質在原子層面的結構，而且水溶液中的結構比起蛋白質晶體更接近於蛋白質在生物體中的自然狀態。目前以自動化的方法來決定 NMR 結構，已經扮演了一個越來越顯著的角色，並且被期待著，大幅地取代傳統以人為的方式來決定水溶液中蛋白質的三度空間結構，並進而加速決定蛋白質結構的速度；以 NMR 決定蛋白質的三度空間結構，所須收集的資訊及步驟如圖 2 所示。

NOE (nuclear Overhauser effect) 訊號的歸屬通常是決定 NMR 結構的瓶頸，雖然共振訊號化學位移的歸屬可以在短時間內完成，但 NOE 訊號的歸屬卻因為有著很多種歸屬的可能性，再加上訊號的

以 CYANA 自動化計算 NMR 蛋白質結構

- 1 共振頻率：依局部環境而定，每一個原子 (nuclear spin) 都能被 NMR 偵測到的獨特頻率。
- 2 透過空間的相互作用：在蛋白質中，氫原子在空間中的距離小於 5 Å (5×10^{-10} m)，會有 NOE (Nuclear Overhauser Effect) 作用。
- 3 cross peaks：鄰近原子間的 NOE 作用在二度、三度或四度空間的 NMR 光譜中所產生的信號 ("peaks")。
- 4 NOESY NMR 光譜：NOESY 光譜提供了蛋白質中數以千計的氫原子間距離的資訊。
- 5 NOE 歸屬 (assignment)：鑑定每一個 cross peak 屬於蛋白質中的哪一個原子對。
- 6 重疊問題：通常一些氫原子會有幾乎相同的共振頻率，重疊的問題，使得以人為方式，把每一個 cross peak 歸屬為單一氫原子對，變得困難而且辛苦。
- 7 自動化 NOE 歸屬：以 CYANA 完全自動化地進行 NOESY cross peak 的歸屬。自動化對於以 NMR 決定蛋白質結構的整體效率是必須的。
- 8 結構計算：從蛋白質中原子之間的距離以及胺基酸的順序，使用 CYANA 計算原子在空間中的排列。
- 9 高效能的計算：結構計算需要大量的計算時間，CYANA 最多能平行地使用 128 個微處理器 (CPU)。
- 10 蛋白質結構：決定 NMR 結構其最後的結果就是蛋白質在水溶液中的三度空間結構。
- 11 結構-功能的關係：蛋白質結構能夠反映出蛋白質的功能，並形成藥物發展的基礎。

圖 2. 以 CYANA 進行自動化計算 NMR 蛋白質結構時，所需的資訊及步驟。

重疊以及雜訊的產生，而變得更加困難，特別是當需要分析有數千個 NOE 訊號的中等分子量蛋白質時，即使是三度空間的光譜，以人為的方式來歸屬數個 NOESY (NOE spectroscopy) 光譜的訊號，通常是耗費時間的 (圖 3)。為了加快決定 NMR 蛋白質結構的速度，一個以電腦為基礎，重複性地同時進行自動化 NOE 訊號歸屬以及結構計算的方法，CYANA (combined assignment and dynamics algorithm for NMR applications)，已被使用來計算蛋白質的結構，此種自動化訊號歸屬/結構計算的方法 (圖 4) 並不需要用到人為的分析，但卻必須克服兩個主要的障礙：第一，能夠被歸屬成單一可能性的 NOE 訊號的數目，通常不足以決定出一蛋白質的摺疊 (fold)，因此，此自動化的方法必須能夠分析那些不能被歸屬成單一可能性的 NOE 訊號；第二，自動化的程式必須能夠處理 NMR 實驗資料

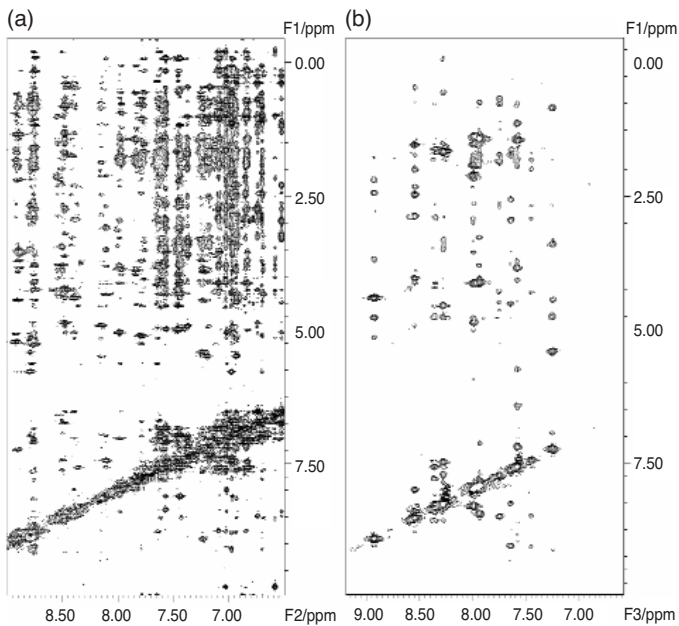


圖 3. Sud (polysulfide-sulfur transferase)⁽⁶⁾ 的 (a) 2D NOESY 光譜 (F1: -0.5–10 ppm、F2: 6.5–9 ppm)；(b) 從 3D ¹⁵N-NOESY 光譜的 128 個平面中取出的一個平面。以人為的方式歸屬每一個 cross peak 將是耗費時間的。

的瑕疵，譬如被錯誤選取的 NOE 訊號，或者 NOE 訊號的位置不準確以及化學位移的歸屬不完全 (典型的實驗數據) 等。自動化的方法比起傳統以人為方式分析的方法，更為快速而且客觀，NOESY 光譜的分析在決定 NMR 蛋白質結構上，不再是一個耗時的步驟。

除了 CYANA 之外，目前也有其他軟體能進行自動化 NOE 歸屬及結構計算，例如：ARIA⁽⁷⁻⁸⁾、KNOWNOE⁽²⁾ 和 AutoStructure⁽⁹⁾ 等，其中廣泛使用的 ARIA 軟體，利用了許多新的觀念，最重要地就是模糊的距離限制 (ambiguous distance constraints) (見以下) 的使用，用以處理最初以化學位移為基礎、進行 NOE 訊號歸屬時所遇到的模糊性；而這些軟體的一般原理，就是把一個光譜學家在歸屬一系列的 NOESY 光譜時所會用到的規則自動化。

二、自動化 NOE 歸屬與結構計算的一般原理

1. NOE 作用

NOE 是 NMR 決定一個蛋白質的二級和四級結構的主要資訊，因為 NOE 訊號代表著在胺基酸殘基中，某氫原子對的距離小於 5 Å (1 Å = 10⁻¹⁰ m)，而這些胺基酸殘基也許在蛋白質的序列上是相隔很遠但在空間上卻是很靠近的。以 NMR 決定蛋白質結構，主要是依靠蛋白質中鄰近氫原子之間的 NOE 作用所推導出許多的距離限制條件⁽¹⁰⁾，而 NOE 的強度，也就是 NOESY 光譜中譜峰的體積，與兩個自旋量子在空間中的距離的負六次方成正比關係，以此為基礎，NOE 通常被當作兩原子之間的距離上限而非當作是精確的距離，因為分子內的運動或者化學交換通常都會減低 NOE 的強度。

2. 以化學位移為基礎的 NOE 訊號歸屬的模糊性

以 *de novo* 的方式用 NMR 決定水溶液中蛋白質三度空間結構，最主要的結構資訊，就是由 NOE 所推導出的距離上限⁽⁴⁾。為了從 NOESY 光譜中得到距離限制條件，NOE 的訊號必須被歸屬，

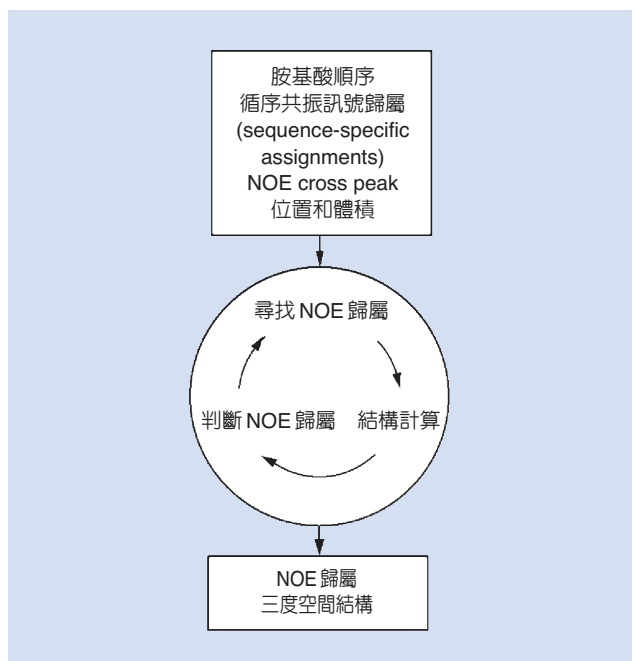


圖 4. 自動化 NOE 歸屬與結構計算的一般流程。

也就是說，互相作用的氫原子對必須被確認。而 NOE 訊號的歸屬主要是由化學位移歸屬所得到的化學位移的值而決定，因為化學位移的值其精確度有限以及訊號的重疊，所以很多 NOE 訊號不能被歸屬為單一的氫原子對，而是由多原子對所組成的不明確歸屬 (ambiguous assignments)。除了訊號的重疊之外，光譜的假訊號 (artifacts)、雜訊、以及因為快速的弛緩 (fast relaxation) 而消失的訊號，這些在收集 NMR 資訊的過程中不可避免的缺點，就是為什麼直到最近，決定蛋白質三度空間結構仍然是一件辛苦而麻煩的工作的主要原因。

3. 模糊的距離限制

模糊的距離限制 (ambiguous distance constraints)⁽⁷⁻⁸⁾ 對於處理以化學位移為基礎的 NOE 訊號歸屬所產生的不明確性，提供了一個非常重要的觀念。在模糊的距離限制被使用之前，一般而言只有被明確歸屬的 NOE (unambiguously assigned NOEs)，被用來當成計算結構的距離限制條件，因為單獨地從化學位移的資訊，大部分的 NOE 並不能被明確地歸屬 (unambiguously assigned)，所以在早期的自動化 NOE 歸屬的演算法中，因缺乏利用模糊距離限制的方法，而大大地阻礙了自動化 NOE 歸屬的功能。當使用模糊距離限制時，每一個 NOE 交叉峰 (cross peak) 都被視為有 n 個訊號的重疊，並且使用一個距離上限作為其有效距離， D_{eff} ：

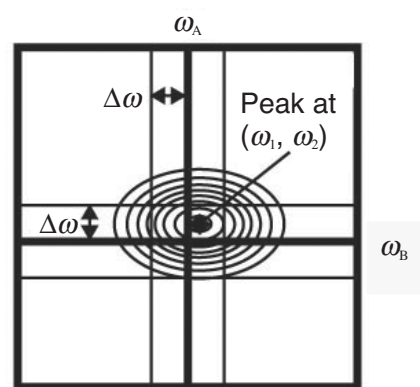
$$D_{\text{eff}} = \left(\sum_{k=1}^n d_k^{-6} \right)^{-1/6}$$

d 為兩氫原子間的距離。相對於一般以人為方式分析的方法尋找明確的訊號歸屬 (unambiguous assignments)，假如 NOE 的訊號不能被明確的歸屬，則不被使用於結構計算之中。而模糊距離限制的觀念，使得可以充分利用 NOESY 光譜中的所有交叉峰，而不論此交叉峰是否有單一的歸屬可能性，因為總是有一些交叉峰包含了一種以上歸屬的可能性，所以模糊距離限制的使用比只用明確的距離限制 (unambiguous distance constraints) 來得實際。

三、CYANA 的自動化 NOE 訊號歸屬／結構計算

1. 概述

在 CYANA 中，自動化的 CANDID⁽¹¹⁻¹²⁾ (combined automated NOE assignment and structure determination module) 演算法重複循環地進行 NOE 訊號歸屬，並使用 CYANA 的扭角動力學演算法 (torsion angle dynamics algorithm) 進行結構計算，在所有的循環當中，都使用相同的輸入資料：蛋白質的胺基酸順序，一組或幾組從循序共振訊號歸屬 (sequence-specific resonance assignments) 所得到的化學位移資料，以及一組或幾組包含交叉峰的位置和體積的二度、三度或四度空間的 NOESY 光譜。在 CANDID 的循環中，每一個 NOE 交叉峰，依據使用者所定義的化學位移範圍 (圖 5)，產生包含許多氫原子對歸屬的可能性，接著對於每一個交叉峰，一些標準被用來判斷這些歸屬的可能性，可能性太低的將被排除；這些標準包括化學位移的值和 NOE 交叉峰的位置是否符合、在全部 NOE 中的一致性，以及與在 CANDID 的循環中所計算出來的蛋白質三度空間結構的相容性。其餘沒有被排除的歸屬 (assignments)，被轉換為距離上限 (upper distance limits)，然後依據距離上限計算出蛋白質



$$|\omega_1 - \omega_A| < \Delta\omega \quad |\omega_2 - \omega_B| < \Delta\omega$$

圖 5. $\Delta\omega$ 為使用者所定義的化學位移範圍； ω_1 、 ω_2 為 NOE 交叉峰的位置； ω_A 、 ω_B 為化學位移的值；在 NOE 的歸屬中， $\Delta\omega$ 必須大於 $|\omega_1 - \omega_A|$ 及 $|\omega_2 - \omega_B|$ 。

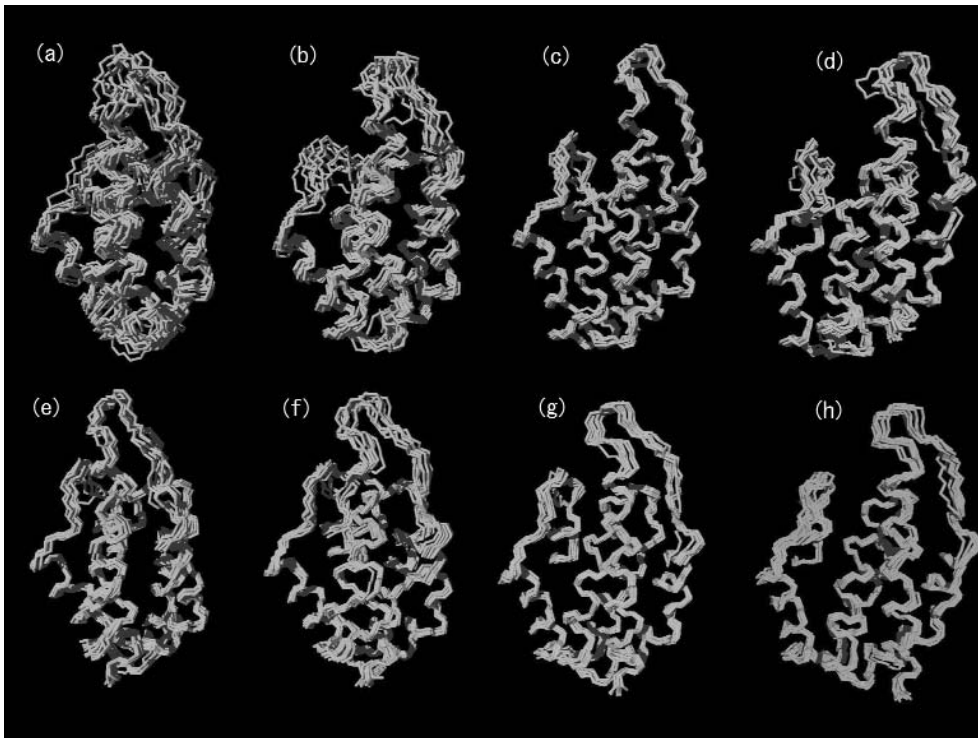


圖 6. 以 CYANA 進行自動化 NOE 歸屬／結構計算時，ENTH-VHS domain At3g16270⁽¹³⁾ 從循環 1 (a) 到循環 7 (g) 的結構以及最後的結構 (h)。

的三度空間結構。以往距離上限都是由只有單一歸屬可能性的交叉峰而來，但模糊的距離限制 (ambiguous distance constraints) 卻包含了數種歸屬的可能性，一個交叉峰如果有太低歸屬的可能性，將被暫時地排除而不被使用。CYANA 的結構計算通常包含了七個循環 (圖 6)，結構的精確性通常隨著每一個循環的增加而改善，所以用來判斷距離限制及歸屬可能性的標準，也隨著循環的增加而更加嚴格。

2. Network-anchoring 和 Constraint combination

Network-anchoring 是指被正確歸屬的距離限制條件，會形成前後一致互相關聯並且足以決定蛋白質三度空間結構的一組資料，因此 network-anchoring 可以判斷 NOE 歸屬的一致性，而不需要蛋白質三度空間結構的知識，也因此彌補了以 *de novo* 決定結構時，缺乏三度空間結構資訊的不足。每一個 NOE 的歸屬必須被加入到所有其他的歸屬中，使得 network-anchoring 成為一個有效的方法來偵測錯誤的距離限制。圖 7 顯示了使用 network-anchoring 之後結構的不同，在使用

network-anchoring 的情形下所得到的結構，明顯地比沒有使用 network-anchoring 的結構收斂，並且已有了正確的折疊 (fold)。

在以 NMR 決定生物分子結構的實際應用上，來自於雜訊及光譜的假訊號 (spectral artifacts) 所產生錯誤的距離限制，也許會被使用在計算結構之中，而 constraint combination 主要在於把這些錯誤的距離限制所帶來的傷害減到最低，它是藉由組合兩個不同且不相關的交叉峰的歸屬而產生，也因此所產生的模糊的距離限制，有較低的錯誤機率。在 CANDID 中，constraint combination 主要用於第一個及第二個循環。在圖 8 中，有使用 constraint combination 的結構，相較於沒有使用 constraint combination 的結構，亦明顯地收斂，而且也呈現出正確的折疊。

3. 不完整的化學位移歸屬的影響

自動化 NOE 歸屬的應用主要是以完整的化學位移資訊為基礎，目前化學位移的歸屬，雖然可望朝向自動化發展⁽⁹⁾，但大部分仍然停留在以互動式或半自動式的方法來歸屬化學位移。經驗顯示，通常大部分的化學位移都可以被歸屬，但仍有一部分

的化學位移是難以歸屬而需要花費更多的時間。因此，假如 NOE 的歸屬及結構的計算，能以不完整的化學位移資料為基礎而進行，那麼以核磁共振決定蛋白質結構的速度將大大地加快。不完整的化學位移歸屬對 NMR 結構可靠性的影響，已經藉由隨機性地省略 ^1H 化學位移而被研究探討⁽¹⁴⁾。以 *Bombyx mori* pheromone binding protein form A (BmPBP^A)⁽¹⁵⁾ 及 *Williopsis mrakii* killer toxin (WmKT)⁽¹⁶⁾ 兩蛋白質為例，當增加省略的比例到約 10% 時，所得的結構僅稍有不同，之後，結構便突然地變得與參考結構不一致，顯示了不正確的結構已經產生；而 CYANA 的 target function 則與省略的比例無關，幾乎總是在小於 5 \AA^2 的範圍；隨著省略比例的增加，沒有歸屬的 NOE 交叉峰的百分比也增加，而最後一個循環的距離限制數目則減少。雖然 BmPBP^A (142 個胺基酸殘基) 比 WmKT (88 個胺基酸殘基) 的分子量大，但是異核 ^{13}C 及 ^{15}N 同位素標定的蛋白質 BmPBP^A 所得的結果，卻比同核無同位素標定的蛋白質 WmKT 所得的結果要好，這是由於 ^{13}C 及 ^{15}N 的化學位移有助於分辨出許多 ^1H 化學位移的退化 (degeneracy)，進而減少了錯誤的 NOE 歸屬的機率。總而言之，大約 90% 的化學位移歸屬是必須的⁽¹⁴⁾。

4. 不完整的 NOESY Peak Picking 的影響

相對於省略化學位移的歸屬，隨機性地省略 NOESY 交叉峰，並不會造成嚴重的問題⁽¹⁴⁾，即使當 50% 的 NOESY 交叉峰被省略，大部分的結構與正確的結構相比較，仍然在 2 \AA 的範圍之內，因此以 CYANA 計算蛋白質結構，允許不完整的 NOESY peak picking，即使最多達 50% 的交叉峰被省略，所得到的結構，其準確性只有稍微減低，這說明了化學位移歸屬的正確性比收集大量的 NOESY 交叉峰來得重要。

四、結論

對於可靠的自動化 NOE 歸屬，大約 90% 的化學位移歸屬是必須的，然而此演算法對於異核的 NOESY 光譜 (heteronuclear-resolved NOESY spectra)

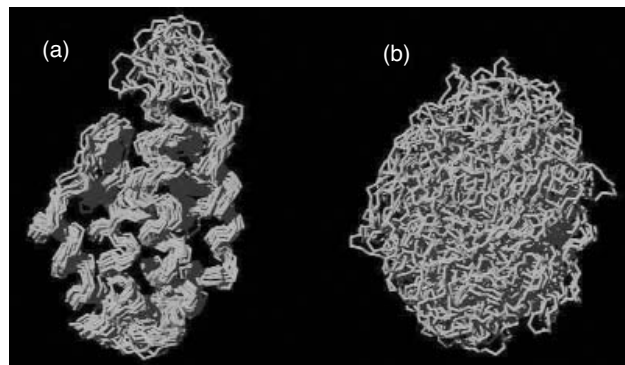


圖 7.(a) 使用 network-anchoring 及 (b) 沒有使用 network-anchoring 時，ENTH-VHS domain At3g16270 在循環 1 的結構。

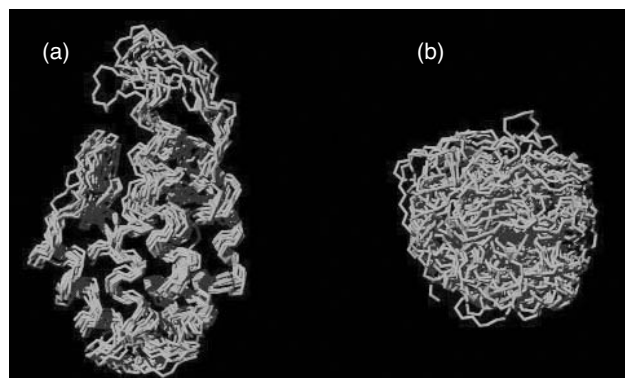


圖 8.(a) 使用 constraint combination 及 (b) 沒有使用 constraint combination 時，ENTH-VHS domain At3g16270 在循環 1 的結構。

卻允許最多 50% 的 NOESY 交叉峰被省略，而不會產生相較於參考結構，RMSD 值大於 2 \AA 的結構；當使用同核的 NOESY 光譜 (homonuclear NOESY spectra) 時，則此演算法會較容易因為資料的不足而受點影響。

以 CYANA 的方法所進行的自動化 NOE 歸屬及蛋白質結構計算，產生類似於以傳統互動式的方法所產生的 NOE 歸屬及蛋白質三度空間結構。在日本理化學研究所的結構基因體／蛋白質體計畫 (RIKEN Structural Genomics/Proteomics Project) 以及其它地方，已使用 CYANA 決定了幾百個以上的蛋白質結構，這些以 *de novo* 的方式所決定的結構，顯示了以 CYANA 的方法所進行的自動化 NOE 歸屬，比傳統以人為方式分析的方法更為快

速且客觀，在以 NMR 決定蛋白質結構的過程中，NOESY 光譜的分析不再是一個時間決定 (time-limiting) 步驟。

誌謝

本文之完成要特別感謝日本理化學研究所基因體科學總合研究中心宮澤辰雄記念計畫對此工作的支持。

參考文獻

1. A. S. Altieri and R. A. Byrd, *Curr. Opin. in Struct. Biol.*, **14**, 547 (2004).
2. W. Gronwald, S. Moussa, R. Elsner, A. Jung, B. Ganslmeier, J. Trenner, W. Kremer, K. P. Neidig, and H. R. Kalbitzer, *J. Biomol. NMR*, **23**, 271 (2002).
3. P. Güntert, Prog. in NMR Spectrosc., **43**, 105 (2003).
4. K. Wüthrich, *NMR of Proteins and Nucleic Acids*, Wiley (1986).
5. J. Cavanagh, W. J. Fairbrother, A. G. Palmer III, and N. J. Skelton, *Protein NMR Spectroscopy* (1996).
6. Y.-J. Lin, F. Dancea, F. Löhr, O. Klimmek, S. Pfeiffer-Marek, M. Nilges, H. Wienk, A. Kröger, and H. Rüterjans, *Biochemistry*, **43**, 1418 (2004).
7. M. Nilges, *Proteins*, **17**, 297 (1993).
8. M. Nilges, *J. Mol. Biol.*, **245**, 645 (1995).
9. M. C. Baran, Y. J. Huang, H. N. Moseley, and G. T. Montelione, *Chem. Rev.*, **104**, 3541 (2004).
10. P. Güntert, *Q. Rev. Biophys.*, **31** (2), 145 (1998).
11. P. Güntert, C. Mumenthaler, and K. Wüthrich, *J. Mol. Biol.*, **273**, 283 (1997).
12. T. Herrmann, P. Güntert, and K. Wüthrich, *J. Mol. Biol.*, **319**, 209 (2002).
13. B. Lopez-Mendez, D. Pantoja-Uceda, T. Tomizawa, S. Koshiba, T. Kigawa, M. Shirouzu, T. Terada, M. Inoue, T. Yabuki, M. Aoki, E. Seki, T. Matsuda, H. Hirota, M. Yoshida, A. Tanaka, T. Osanai, M. Seki, K. Shinozaki, S. Yokoyama, and P. Güntert, *J. Biomol. NMR*, **29**, 205 (2004).
14. J. G. Jee and P. Güntert, *J. Struct. Funct. Genom.*, **4**, 179 (2003).
15. R. Horst, F. Damberger, P. Luginbuhl, P. Güntert, G. Peng, L. Nikonova, W. S. Leal, and K. Wüthrich, *Proc. Natl. Acad. Sci. USA* **98**, 14374 (2001).
16. W. Antuch, P. Güntert, and K. Wüthrich, *Nat. Struct. Biol.*, **3**, 662 (1996).

- 林意然先生為德國法蘭克福歌德大學生物化學博士，現任日本理化學研究所基因體科學總合研究中心副研究員。
- Peter Güntert 先生為瑞士蘇黎世聯邦技術學院物理博士，現任日本理化學研究所基因體科學總合研究中心宮澤辰雄記念計畫上級研究員。
- Yi-Jan Lin received his Ph.D. in biochemistry from the Johann Wolfgang Goethe-University of Frankfurt, Germany. He is currently a research associate of RIKEN Genomic Sciences Center, Japan.
- Peter Güntert received his Ph.D. in physics from the Swiss Federal Institute of Technology (ETH), Zurich, Switzerland. He is currently a principal investigator of the Tatsuo Miyazawa Memorial Program of RIKEN Genomic Sciences Center, Japan.